

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2002, 15:59:31 ; Search time 31 Seconds
(without alignments)
85.968 Million cell updates/sec

Title: US-09-674-738-1
Perfect score: 108
Sequence: 1 VPOSIDWRDYGAVTSVKNON 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
A.GeneSeq_101002.*
1: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	108	100.0	20	AAV51581	Pineapple bromelain
2	108	100.0	213	AAV47550	Stem bromelain pro
3	108	100.0	213	AAV48974	Pineapple stem pro
4	99	91.7	311	AAV83105	710 proteainase fto
5	99	91.7	324	AAV83107	D18548 proteainase
6	99	91.7	351	AAV83104	674 proteainase fto
7	97	89.8	20	AAV80802	Comosain N-termina
8	95	88.0	186	AAV83098	569 proteainase fto
9	95	88.0	186	AAV83101	573 proteainase fto
10	95	88.0	187	AAV83099	571 proteainase fto

11	95	88.0	188	AAV83100	572 proteainase fto
12	95	88.0	150	AAV83106	D1403 proteainase
13	93	86.1	20	AAV80801	Ananain N-terminal
14	93	86.1	20	AAV51582	Pineapple bromelain
15	93	86.1	21	AAV84395	N-terminal peptide
16	93	86.1	21	AAV84389	N-terminal peptide
17	93	86.1	28	AAV84391	N-terminal peptide
18	88	81.5	15	AAV84391	Bromelain protein
19	88	81.5	18	AAV84394	N-terminal peptide
20	82	75.9	15	AAV84392	N-terminal peptide
21	79	73.1	127	AAV84392	Cysteine protease
22	79	73.1	148	AAV84392	Cysteine protease
23	79	73.1	258	AAV84392	Arbidopsin thalia
24	79	73.1	258	AAV84392	Cysteine protease
25	79	73.1	278	AAV84392	Cysteine protease
26	79	73.1	297	AAV84392	Arbidopsin thalia
27	78	72.2	300	AAV84290	Arbidopsin thalia
28	78	72.2	324	AAV84291	Arbidopsin thalia
29	78	72.2	361	AAV84291	Arbidopsin thalia
30	78	72.2	362	AAV84291	Arbidopsin thalia
31	78	72.2	368	AAV84291	Arbidopsin thalia
32	77	71.3	15	AAV84388	N-terminal peptide
33	77	71.3	15	AAV84388	N-terminal peptide
34	75	69.4	21	AAV84396	N-terminal peptide
35	74	68.5	277	AAV84396	Arbidopsin thalia
36	74	68.5	277	AAV84396	Arbidopsin thalia
37	74	68.5	277	AAV84396	Arbidopsin thalia
38	73	67.6	15	AAV84397	N-terminal peptide
39	73	67.6	20	AAV84397	N-terminal peptide
40	73	67.6	21	AAV84397	N-terminal peptide
41	73	67.6	21	AAV84397	N-terminal peptide
42	72	66.7	396	AAV84397	Cysteine protease
43	72	66.7	129	AAV84397	Cysteine protease
44	72	66.7	191	AAV84397	Arbidopsin thalia
45	72	66.7	220	AAV84397	Arbidopsin thalia

ALIGNMENTS

RESULT 1	AAV51581	standard; peptide; 20 AA.
ID	AAV51581	
XX	AAV51581	
XX	18-MAY-2000	(first entry)
XX		
DE	Pineapple bromelain protease peptide fragment from fraction 4.	
XX		
KW	Bromelain; pineapple; protease; blood coagulation; inhibitor; plasmin.	
XX		
OS	Ananas comosus.	
XX		
PN	WO200003729-A1.	
XX		
PD	27-JAN-2000.	
XX		
PF	15-JUL-1998; 98WO-EP04406.	
XX		
PR	15-JUL-1998; 98WO-EP04406.	
XX		
PA	(URSA-) UNISAPPHARM ARZNEIMITTEL GMBH.	
XX		
PI	Maurer R, Eckert K, Grabowska E, Eschmann K;	
XX		
DR	WPI; 2000-182338/16.	
XX		
PT	Composition for inhibiting blood coagulation more efficiently than	
XX	crude bromelain powder comprises bromelain proteases -	
PS	Claim 6, Page 15, 24pp; German.	
XX		

CC This invention describes a novel use of bromelain proteases (I) for
 CC inhibiting blood coagulation. (I) stimulate formation of plasmin, but
 CC inhibit both formation of fibrin and adhesion of thrombocytes to
 CC endothelial cells. (I) are about 10 times more active than bromelain
 CC crude powder, i.e. comparable with streptokinase. AAY51581 and AAY51582
 CC represent bromelain proteases isolated from pineapple (Ananas comosus)
 CC which are used in the method of the invention.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 108; DB 21; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.1e-11;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKQN 20

DB 1 VPOSIDWRDYGAVTSVKQN 20

RESULT 2

AAR47550
 ID AAR47550 standard; protein; 213 AA.

XX AC AAR47550;

XX DT 18-JUL-1994 (first entry)

XX DE Stem bromelain protease.

XX KW Proteolytic enzyme; cysteine protease; diarrhoea; fluid secretion;
 KW gastro-intestinal tract; GI tract; absorption.

XX OS Ananas comosus.

XX PN WO9400147-A.

XX PD 06-JAN-1994.

XX PF 30-JUN-1993; 93WO-GB01374.

XX PR 30-JUN-1992; 92GB-0013862.

XX PR 20-APR-1993; 93GB-0008184.

XX PR 25-JUN-1993; 93GB-0013189.

XX FA (CORT-) CORTECS LTD.

XX PI Mynott TL;

XX DR WPI; 1994-025889/03.

XX PT Enzyme, esp. bromelain, contg. compsn. for control or diarrhoea -
 PT esp. when caused by heat-labile toxin-producing E. coli. also

XX PT inhibits fluid secretion in the gut and promotes absorption

XX PS Disclosure; Page 11-13; 70pp; English.

XX CC The sequence is that of stem bromelain protease, a mixture of
 CC various moieties derived from the stem of the pineapple. The
 CC protein is a cysteine protease used to treat infectious diarrhoea
 CC caused by heat labile or heat stable toxins, enterotoxigenic E. coli
 CC and Vibrio cholerae. The enzyme reduces toxin binding activity,
 CC inhibits the secretory effect of the toxins and prevents attachment
 CC of bacteria to the intestinal lining. It also inhibits secretion
 CC caused by secretagogues, e.g. PGE2, theophylline, etc. and has no
 CC adverse effect on nutrient influx.

XX SQ Sequence 213 AA;

Query Match 100.0%; Score 108; DB 15; Length 213;

Best Local Similarity 100.0%; Pred. No. 1.8e-10;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKQN 20

DB 2 VPOSIDWRDYGAVTSVKQN 21

RESULT 3

AAR68974
 ID AAR68974 standard; protein; 213 AA.

XX AC AAR68974;

XX DT 09-AUG-1995 (first entry)

XX DE Pineapple stem bromelain protease.

XX KW Bromelain; protease; proteolytic enzyme; cholera; diphtheria; cancer;
 KW inflammation.

XX OS Ananas comosus.

XX PN WO9500169-A.

XX PD 05-JAN-1995.

XX PF 24-JUN-1994; 94WO-GB01368.

XX PR 25-JUN-1993; 93GB-0013188.

XX PA (CORT-) CORTECS LTD.

XX PI Mynott TL;

XX DR WPI; 1995-051751/07.

XX PT Use of purified stem bromelain protease - to treat conditions
 PT dependent on cyclic nucleotide pathways e.g. cholera, diphtheria
 PT or cancer

XX PS Claim 1; Page 35; 45pp; English.

XX CC Pineapple stem bromelain protease is used in the treatment or
 CC prophylaxis of conditions which are dependent on cyclic nucleotide
 CC and other intracellular and/or extracellular pathways. The diseases
 CC include cancer, inflammation, atherosclerosis and bacterial
 CC infections such as cholera, diphtheria and pertussis.

XX SQ Sequence 213 AA;

Query Match 100.0%; Score 108; DB 16; Length 213;

Best Local Similarity 100.0%; Pred. No. 1.8e-10;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKQN 20

DB 2 VPOSIDWRDYGAVTSVKQN 21

RESULT 4

AAY83105
 ID AAY83105 standard; Protein; 311 AA.

XX AC AAY83105;

XX DT 04-JUL-2000 (first entry)

XX DE 710 proteinase from pineapple stem.

XX KW Proteinase; pineapple; prevention; prophylaxis; tumour;
 KW solid tumour; non-solid tumour; virus induced tumour;

XX KW immunostimulant; immunodeficiency; HIV;

XX KW human immunodeficiency virus; malaria; malnutrition; trauma; burn;

XX KW surgery; genetic abnormality; diabetes; nitric oxide; NO;

XX KW antimicrobial.

OS Ananas comosus.
 PN WO200014253-A1.
 XX 16-MAR-2000.
 XX 24-AUG-1999; 99WO-GB02784.
 PF 02-SEP-1998; 98GB-0019138.
 XX (CORT-) CORTECS UK LTD.
 PA Mynott TL, Crossett B;
 PI WPI; 2000-256998/22.
 DR N-PSDB; AA293389.
 XX Stem bromelain-like proteinases useful for the prevention and treatment
 PT of cancer and diseases or conditions which respond to increased nitric
 PT oxide production are encoded by nucleic acid sequences isolated from a
 PT pineapple stem -
 PS Claim 4; Page 58-59; 72pp; English.
 XX The sequences encoding proteinases from pineapple stem have
 CC the same activity as a stem bromelain-like proteinase, CCX2.
 CC The proteinases can be used in human or veterinary medicine and
 CC can be used for treating or preventing cancer, e.g. solid tumours;
 CC non-solid tumours and virus induced tumours. The proteinases can
 CC also be used as an immunostimulant or in the preparation of an
 CC immunostimulant to treat immunodeficiencies arising from
 CC malnutrition, infection (e.g. HIV and malaria), tumours (e.g.
 CC lymphoid, myeloma), trauma (e.g. burns, wounds, surgery) medical
 CC treatment (e.g. with drugs such as steroids, cyclosporin and
 CC cyclophosphamide), protein loss (e.g. in diarrhoea and burns),
 CC genetic abnormalities (e.g. in combined immunodeficiency patients
 CC lacking B or T cells), diabetes and old age. The proteinase can
 CC also be used in the treatment or prevention of diseases or
 CC conditions which respond to increased nitric oxide (NO) production,
 CC or as an antimicrobial agent or the component of an antimicrobial
 CC agent.
 CC Sequence 311 AA;
 SQ
 Query Match 91.7%; Score 99; DB 21; Length 311;
 Best Local Similarity 90.0%; Pred. No. 1e-08; 2; Indels 0; Gaps 0;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 VPOSIDWRDYGAVTSYKNON 20
 DB 96 VPOSIDWRDYGAVTSYKNON 115
 RESULT 5
 ID AAY83107 standard; Protein; 324 AA.
 AC AAY83107;
 XX 04-JUL-2000 (first entry)
 DE D385848 proteinase from pineapple stem.
 KW Proteinase; pineapple; treatment; prevention; prophylaxis; tumour;
 KW solid tumour; non-solid tumour; virus induced tumour;
 KW immunostimulant; immunodeficiency; HIV;
 KW human immunodeficiency virus; malaria; malnutrition; trauma; burn;
 KW surgery; genetic abnormality; diabetes; nitric oxide; NO;
 KW antimicrobial.
 OS Ananas comosus.
 XX WO200014253-A1.
 PN 16-MAR-2000.

XX 16-MAR-2000.
 PD 24-AUG-1999; 99WO-GB02784.
 PF 02-SEP-1998; 98GB-0019138.
 XX (CORT-) CORTECS UK LTD.
 PA Mynott TL, Crossett B;
 PI WPI; 2000-256998/22.
 DR N-PSDB; AA293391.
 XX Stem bromelain-like proteinases useful for the prevention and treatment
 PT of cancer and diseases or conditions which respond to increased nitric
 PT oxide production are encoded by nucleic acid sequences isolated from a
 PT pineapple stem -
 PS Claim 4; Page 64-66; 72pp; English.
 XX The sequences encoding proteinases from pineapple stem have
 CC the same activity as a stem bromelain-like proteinase, CCX2.
 CC The proteinases can be used in human or veterinary medicine and
 CC can be used for treating or preventing cancer, e.g. solid tumours;
 CC non-solid tumours and virus induced tumours. The proteinases can
 CC also be used as an immunostimulant or in the preparation of an
 CC immunostimulant to treat immunodeficiencies arising from
 CC malnutrition, infection (e.g. HIV and malaria), tumours (e.g.
 CC lymphoid, myeloma), trauma (e.g. burns, wounds, surgery) medical
 CC treatment (e.g. with drugs such as steroids, cyclosporin and
 CC cyclophosphamide), protein loss (e.g. in diarrhoea and burns),
 CC genetic abnormalities (e.g. in combined immunodeficiency patients
 CC lacking B or T cells), diabetes and old age. The proteinase can
 CC also be used in the treatment or prevention of diseases or
 CC conditions which respond to increased nitric oxide (NO) production,
 CC or as an antimicrobial agent or the component of an antimicrobial
 CC agent.
 CC Sequence 324 AA;
 SQ
 Query Match 91.7%; Score 99; DB 21; Length 324;
 Best Local Similarity 90.0%; Pred. No. 1e-08; 2; Indels 0; Gaps 0;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 VPOSIDWRDYGAVTSYKNON 20
 DB 96 VPOSIDWRDYGAVTSYKNON 115
 RESULT 6
 ID AAY83104 standard; Protein; 351 AA.
 AC AAY83104;
 XX 04-JUL-2000 (first entry)
 DE 674 proteinase from pineapple stem.
 KW Proteinase; pineapple; treatment; prevention; prophylaxis; tumour;
 KW solid tumour; non-solid tumour; virus induced tumour;
 KW immunostimulant; immunodeficiency; HIV;
 KW human immunodeficiency virus; malaria; malnutrition; trauma; burn;
 KW surgery; genetic abnormality; diabetes; nitric oxide; NO;
 KW antimicrobial.
 OS Ananas comosus.
 XX WO200014253-A1.
 PN 16-MAR-2000.

PF 24-AUG-1999; 99WO-GB02784.
 XX
 PR 02-SEP-1998; 98GB-0019138.
 XX
 PA (CORT-) CORTECS UK LTD.
 XX
 PI Mynott TL, Crossett B;
 XX
 XX WPI; 2000-256998/22.
 DR
 DR N-PSDB; AA293388.
 XX
 XX
 PT Stem bromelain-like proteinases useful for the prevention and treatment
 PT of cancer and diseases or conditions which respond to increased nitric
 PT oxide production are encoded by nucleic acid sequences isolated from a
 PT pineapple stem -
 XX
 XX Claim 4; Page 55-57; 72pp; English.
 PS
 CC The sequences encoding proteinases from pineapple stem have
 CC the same activity as a stem bromelain-like proteinase, CCX2.
 CC The proteinases can be used in human or veterinary medicine and
 CC can be used for treating or preventing cancer, e.g. solid tumours;
 CC non-solid tumours and virus induced tumours. The proteinases can
 CC also be used as an immunostimulant or in the preparation of an
 CC immunostimulant to treat immunodeficiencies arising from
 CC malnutrition, infection (e.g. HIV and malaria), tumours (e.g.
 CC lymphoid, myeloma), trauma (e.g. burns, wounds, surgery) medical
 CC treatment (e.g. with drugs such as steroids, cyclosporin and
 CC cyclophosphamide), protein loss (e.g. in diarrhoea and burns),
 CC genetic abnormalities (e.g. in combined immunodeficiency patients
 CC lacking B or T cells), diabetes and old age. The proteinase can
 CC also be used in the treatment or prevention of diseases or
 CC conditions which respond to increased nitric oxide (NO) production,
 CC or as an antimicrobial agent or the component of an antimicrobial
 CC agent.
 XX
 XX Sequence 351 AA;
 SQ
 Query Match 91.7%; Score 99; DB 21; Length 351;
 Best Local Similarity 90.0%; Pred. No. 1.1e-08;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 VPOSIDWRDYGAVTSVKQN 20
 |||||
 DB 123 VPOSIDWRDYGAVNEVKQN 142
 |||||
 RESULT 7
 AA080802
 ID AA080802 standard; peptide; 20 AA.
 XX
 AC AA080802;
 XX
 DT 24-NOV-1998 (first entry)
 XX
 DE Comosain N-terminal sequence.
 XX
 KW Comosain; bromelain; CCS; autoimmune; immunosuppressant;
 KW Graft rejection; antiallergic; antiapoptosis; toxic shock;
 KW antiparasitic; antipathogenic; anticancer; antitumour.
 XX
 OS Ananas comosus.
 XX
 PN WO9838291-A1.
 XX
 PD 03-SEP-1998.
 XX
 PF 25-FEB-1998; 98WO-GB00590.
 XX
 PR 25-MAR-1997; 97GB-0006119.
 PR 25-FEB-1997; 97GB-0003827.
 PR 25-FEB-1997; 97GB-0003850.
 PR 28-FEB-1997; 97GB-0004252.

XX (CORT-) CORTECS UK LTD.
 PA Engwerda C, Mynott TL, Peek K;
 PI
 XX WPI; 1998-481194/41.
 DR
 XX
 XX New isolated component of bromelain - used for treating e.g.
 PT autoimmune diseases, transplant rejection, allergic reactions,
 PT toxic shock, apoptosis, parasite or pathogen infections or cancer
 XX
 XX Disclosure; Page 24; 56pp; English.
 PS
 CC The patent discloses a fraction of bromelain, designated CCS, which
 CC (i) contains proteins having molecular weights of 15.07 kD, 25.85 kD
 CC and 27.45 kD (as determined by SDS-PAGE), (ii) has isoelectric points
 CC of 10.4 and 10.45, and (iii) is obtainable by: (a) dissolving bromelain
 CC in acetate buffer at pH 5.0; (b) separating the components by fast flow
 CC high performance chromatography on 8-Sepharose (RPM), eluting with a
 CC linear gradient of 0 to 0.8M NaCl in acetate buffer over 300 ml; (c)
 CC collecting the fraction corresponding to the final double peak off the
 CC column; and (d) isolating the protein from the fraction collected in (c).
 CC The CCS fraction can be used: (a) for modulating intracellular signalling
 CC pathways which control cell growth and proliferation; (b) for inhibiting
 CC the production of growth factors and cytokines by cells; (c) for reducing
 CC or preventing the activation of mitogen-activated protein (MAP) kinase
 CC pathway; (d) for reducing or preventing the activation of T cells;
 CC (e) as an immunosuppressive agent; (f) for blocking the production of
 CC growth factors and other cytokines; or (g) for the treatment or
 CC prevention of autoimmune diseases, graft or transplant rejection,
 CC allergic reactions, toxic shock, apoptosis, parasite or pathogen
 CC infections, or cancer. The CCS fraction is postulated to consist of
 CC the cysteine proteases ananain and comosain, together with various other
 CC components. The present sequence represents the N-terminal of comosain.
 XX
 XX Sequence 20 AA;
 SQ
 Query Match 89.8%; Score 97; DB 19; Length 20;
 Best Local Similarity 94.7%; Pred. No. 8.8e-10;
 Matches 18; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VPOSIDWRDYGAVTSVKQN 19
 |||||
 DB 1 VPOSIDWRNYGAVTSVKQN 19
 |||||
 RESULT 8
 AA083098
 ID AA083098 standard; Protein; 186 AA.
 XX
 AC AA083098;
 XX
 DT 04-JUL-2000 (first entry)
 XX
 DE 569 proteinase from pineapple stem.
 XX
 KW Proteinase; pineapple; treatment; prevention; prophylaxis; tumour;
 KW solid tumour; non-solid tumour; virus induced tumour;
 KW immunostimulant; immunodeficiency; HIV;
 KW human immunodeficiency virus; malaria; malnutrition; trauma; burn;
 KW surgery; genetic abnormality; diabetes; nitric oxide; NO;
 KW antimicrobial.
 XX
 OS Ananas comosus.
 XX
 PN WO200014253-A1.
 XX
 PD 16-MAR-2000.
 XX
 PF 24-AUG-1999; 99WO-GB02784.
 XX
 PR 02-SEP-1998; 98GB-0019138.
 XX

PA (CORT-) CORTECS UK LTD.
 XX
 XX Mynote TL, Crossett B;
 PI
 XX
 DR WPI; 2000-256998/22.
 N-PSDB; AA293382.
 XX
 PT Stem bromelain-like proteinases useful for the prevention and treatment
 of cancer and diseases or conditions which respond to increased nitric
 oxide production are encoded by nucleic acid sequences isolated from a
 pineapple stem -
 PT
 XX
 PS Claim 4; Page 44-45; 72pp; English.
 XX
 XX The sequences encoding proteinases from pineapple stem have
 CC the same activity as a stem bromelain-like proteinase, CCX2.
 CC The proteinases can be used in human or veterinary medicine and
 CC can be used for treating or preventing cancer, e.g. solid tumours;
 CC non-solid tumours and virus induced tumours. The proteinases can
 CC also be used as an immunostimulant or in the preparation of an
 CC immunostimulant to treat immunodeficiencies arising from
 CC malnutrition, infection (e.g. HIV and malaria), tumours (e.g.
 CC lymphoid, myeloma), trauma (e.g. burns, wounds, surgery), medical
 CC treatment (e.g. with drugs such as steroids, cyclosporin and
 CC cyclophosphamide), protein loss (e.g. in diarrhoea and burns),
 CC genetic abnormalities (e.g. in combined immunodeficiency patients
 CC lacking B or T cells), diabetes and old age. The proteinase can
 CC also be used in the treatment or prevention of diseases or
 CC conditions which respond to increased nitric oxide (NO) production,
 CC or as an antimicrobial agent or the component of an antimicrobial
 CC agent.
 CC
 XX
 SQ Sequence 186 AA;
 XX
 Query Match 88.0%; Score 95; DB 21; Length 186;
 Best Local Similarity 89.5%; Pred. No. 2.7e-08;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 POSIDWRDYGAVTSYKXN 20
 DB 1 POSIDWRDYGAVNEKXN 19
 XX
 RESULT 9
 AAY83101
 ID AAY83101 standard; Protein: 186 AA.
 XX
 AC AAY83101;
 XX
 DT 04-JUL-2000 (first entry)
 XX
 DE 573 proteinase from pineapple stem.
 XX
 KW Proteinase; pineapple; treatment; prevention; prophylaxis; tumour;
 KW solid tumour; non-solid tumour; virus induced tumour;
 KW immunostimulant; immunodeficiency; HIV;
 KW human immunodeficiency virus; malaria; malnutrition; trauma; burn;
 KW surgery; genetic abnormality; diabetes; nitric oxide; NO;
 KW antimicrobial.
 XX
 OS Ananas comosus.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 20 /note="Unidentified amino acid"
 FT
 XX
 PN WO200014253-A1.
 XX
 PD 16-MAR-2000.
 XX
 PF 24-AUG-1999; 99WO-GB02784.
 XX
 PR 02-SEP-1998; 98GB-0019138.
 XX
 PA (CORT-) CORTECS UK LTD.

XX
 PA (CORT-) CORTECS UK LTD.
 XX
 XX Mynote TL, Crossett B;
 PI
 XX
 DR WPI; 2000-256998/22.
 N-PSDB; AA293385.
 XX
 PT Stem bromelain-like proteinases useful for the prevention and treatment
 of cancer and diseases or conditions which respond to increased nitric
 oxide production are encoded by nucleic acid sequences isolated from a
 pineapple stem -
 PT
 XX
 PS Claim 4; Page 50-51; 72pp; English.
 XX
 XX The sequences encoding proteinases from pineapple stem have
 CC the same activity as a stem bromelain-like proteinase, CCX2.
 CC The proteinases can be used in human or veterinary medicine and
 CC can be used for treating or preventing cancer, e.g. solid tumours;
 CC non-solid tumours and virus induced tumours. The proteinases can
 CC also be used as an immunostimulant or in the preparation of an
 CC immunostimulant to treat immunodeficiencies arising from
 CC malnutrition, infection (e.g. HIV and malaria), tumours (e.g.
 CC lymphoid, myeloma), trauma (e.g. burns, wounds, surgery), medical
 CC treatment (e.g. with drugs such as steroids, cyclosporin and
 CC cyclophosphamide), protein loss (e.g. in diarrhoea and burns),
 CC genetic abnormalities (e.g. in combined immunodeficiency patients
 CC lacking B or T cells), diabetes and old age. The proteinase can
 CC also be used in the treatment or prevention of diseases or
 CC conditions which respond to increased nitric oxide (NO) production,
 CC or as an antimicrobial agent or the component of an antimicrobial
 CC agent.
 CC
 XX
 SQ Sequence 186 AA;
 XX
 Query Match 88.0%; Score 95; DB 21; Length 186;
 Best Local Similarity 89.5%; Pred. No. 2.7e-08;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 POSIDWRDYGAVTSYKXN 20
 DB 1 POSIDWRDYGAVNEKXN 19
 XX
 RESULT 10
 AAY83099
 ID AAY83099 standard; Protein: 187 AA.
 XX
 AC AAY83099;
 XX
 DT 04-JUL-2000 (first entry)
 XX
 DE 571 proteinase from pineapple stem.
 XX
 KW Proteinase; pineapple; treatment; prevention; prophylaxis; tumour;
 KW solid tumour; non-solid tumour; virus induced tumour;
 KW immunostimulant; immunodeficiency; HIV;
 KW human immunodeficiency virus; malaria; malnutrition; trauma; burn;
 KW surgery; genetic abnormality; diabetes; nitric oxide; NO;
 KW antimicrobial.
 XX
 OS Ananas comosus.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 20 /note="Unidentified amino acid"
 FT
 XX
 PN WO200014253-A1.
 XX
 PD 16-MAR-2000.
 XX
 PF 24-AUG-1999; 99WO-GB02784.
 XX
 PR 02-SEP-1998; 98GB-0019138.
 XX
 PA (CORT-) CORTECS UK LTD.

PI Mynott TL, Crossett B;

XX WPI; 2000-256998/22.

DR N-PSDB; AAZ93383.

XX

PT Stem bromelain-like proteinases useful for the prevention and treatment
of cancer and diseases or conditions which respond to increased nitric
oxide production are encoded by nucleic acid sequences isolated from a
pineapple stem -

XX

PS Claim 4; Page 46-47; 72pp; English.

XX

CC The sequences encoding proteinases from pineapple stem have
the same activity as a stem bromelain-like proteinase, CCX2.
The proteinases can be used in human or veterinary medicine and
can be used for treating or preventing cancer, e.g. solid tumours;
non-solid tumours and virus induced tumours. The proteinases can
also be used as an immunostimulant or in the preparation of an
immunostimulant to treat immunodeficiencies arising from
malnutrition, infection (e.g. HIV and malaria), tumours (e.g.
lymphoid, myeloma), trauma (e.g. burns, wounds, surgery) medical
treatment (e.g. with drugs such as steroids, cyclosporin and
cyclophosphamide), protein loss (e.g. in diarrhoea and burns),
genetic abnormalities (e.g. in combined immunodeficiency patients
lacking B or T cells), diabetes and old age. The proteinase can
also be used in the treatment or prevention of diseases or
conditions which respond to increased nitric oxide (NO) production,
or as an antimicrobial agent or the component of an antimicrobial
agent.

XX

SQ Sequence 187 AA;

Query Match 88.0%; Score 95; DB 21; Length 187;
Best Local Similarity 89.5%; Pred. No. 2.7e-08;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 POSIDWRDYGAVTSVKKNQ 20

DB 1 POSIDWRDYGAVNEVKQN 19

RESULT 11

AAV83100
ID AAV83100 standard; Protein; 188 AA.

XX
AC AAV83100;

XX
DT 04-JUL-2000 (first entry)

XX
DE 572 proteinase from pineapple stem.

XX Proteinase; pineapple; treatment; prevention; prophylaxis; tumour;
solid tumour; non-solid tumour; virus induced tumour;
immunostimulant; immunodeficiency; HIV;
human immunodeficiency virus; malaria; malnutrition; trauma; burn;
surgery; genetic abnormality; diabetes; nitric oxide; NO;
antimicrobial.

XX
OS Ananas comosus.

XX
FH Key Location/Qualifiers
FT Misc-difference 72
FT /note= "Unidentified amino acid"

XX
PN WO200014253-A1.

XX
PD 16-MAR-2000.

XX
PF 24-AUG-1999; 99WO-GB02784.

XX
PR 02-SEP-1998; 98GB-0019138.

XX
PA (CORT-) CORTECS UK LTD.

XX

PI Mynott TL, Crossett B;

XX WPI; 2000-256998/22.

DR N-PSDB; AAZ93384.

XX

PT Stem bromelain-like proteinases useful for the prevention and treatment
of cancer and diseases or conditions which respond to increased nitric
oxide production are encoded by nucleic acid sequences isolated from a
pineapple stem -

XX

PS Claim 4; Page 48-49; 72pp; English.

XX

CC The sequences encoding proteinases from pineapple stem have
the same activity as a stem bromelain-like proteinase, CCX2.
The proteinases can be used in human or veterinary medicine and
can be used for treating or preventing cancer, e.g. solid tumours;
non-solid tumours and virus induced tumours. The proteinases can
also be used as an immunostimulant or in the preparation of an
immunostimulant to treat immunodeficiencies arising from
malnutrition, infection (e.g. HIV and malaria), tumours (e.g.
lymphoid, myeloma), trauma (e.g. burns, wounds, surgery) medical
treatment (e.g. with drugs such as steroids, cyclosporin and
cyclophosphamide), protein loss (e.g. in diarrhoea and burns),
genetic abnormalities (e.g. in combined immunodeficiency patients
lacking B or T cells), diabetes and old age. The proteinase can
also be used in the treatment or prevention of diseases or
conditions which respond to increased nitric oxide (NO) production,
or as an antimicrobial agent or the component of an antimicrobial
agent.

XX

SQ Sequence 188 AA;

Query Match 88.0%; Score 95; DB 21; Length 188;
Best Local Similarity 89.5%; Pred. No. 2.7e-08;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 POSIDWRDYGAVTSVKKNQ 20

DB 1 POSIDWRDYGAVNEVKQN 19

RESULT 12

AAV83106

ID AAV83106 standard; Protein; 190 AA.

XX

AC AAV83106;

XX

DT 04-JUL-2000 (first entry)

XX

DE D14059 proteinase from pineapple stem.

XX

XX Proteinase; pineapple; treatment; prevention; prophylaxis; tumour;
solid tumour; non-solid tumour; virus induced tumour;
immunostimulant; immunodeficiency; HIV;
human immunodeficiency virus; malaria; malnutrition; trauma; burn;
surgery; genetic abnormality; diabetes; nitric oxide; NO;
antimicrobial.

XX

OS Ananas comosus.

XX

PN WO200014253-A1.

XX

PD 16-MAR-2000.

XX

PF 24-AUG-1999; 99WO-GB02784.

XX

PR 02-SEP-1998; 98GB-0019138.

XX

PA (CORT-) CORTECS UK LTD.

XX

PI Mynott TL, Crossett B;

XX

DR WPI; 2000-256998/22.
PT N-PSDB; AA293390.
XX Stem bromelain-like proteinases useful for the prevention and treatment
PT of cancer and diseases or conditions which respond to increased nitric
PT oxide production are encoded by nucleic acid sequences isolated from a
PT pineapple stem -
XX
XX Claim 4; Page 61; 72pp; English.
XX
CC The sequences encoding proteinases from pineapple stem have
CC the same activity as a stem bromelain-like proteinase, CCX2.
CC The proteinases can be used in human or veterinary medicine and
CC can be used for treating or preventing cancer, e.g. solid tumours;
CC non-solid tumours and virus induced tumours. The proteinases can
CC also be used as an immunostimulant or in the preparation of an
CC immunostimulant to treat immunodeficiencies arising from
CC malnutrition, infection (e.g. HIV and malaria), tumours (e.g.
CC lymphoid, myeloma), trauma (e.g. burns, wounds, surgery) medical
CC treatment (e.g. with drugs such as steroids, cyclosporin and
CC cyclophosphamide); protein loss (e.g. in diarrhoea and burns),
CC genetic abnormalities (e.g. in combined immunodeficiency patients
CC lacking B or T cells), diabetes and old age. The proteinase can
CC also be used in the treatment or prevention of diseases or
CC conditions which respond to increased nitric oxide (NO) production,
CC or as an antimicrobial agent or the component of an antimicrobial
CC agent.
XX
XX Sequence 190 AA;
SQ
Query Match 88.0%; Score 95; DB 21; Length 190;
Best Local Similarity 89.5%; Pred. No. 2.8e-08;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 2 POSIDMRDYGAVTSVKNO 20
DB 1 POSIDMRDYGAVTSVKNO 19
RESULT 13
AAW80801
ID AAW80801 standard; peptide; 20 AA.
XX
XX AAW80801;
XX
XX 24-NOV-1998 (first entry)
XX
XX Ananain N-terminal sequence.
XX
XX Ananain; bromelain; CCS; autoimmune; immunosuppressant;
XX graft rejection; antiallergic; antipathogenic; anticancer; antitumour;
XX antiparasitic; antipathogenic; anticancer; antitumour.
XX
XX Ananas comosus.
XX
XX W09838291-A1.
XX
XX 03-SEP-1998.
XX
XX 25-FEB-1998; 98WO-G800590.
XX
XX 25-MAR-1997; 97GB-0006119.
XX 25-FEB-1997; 97GB-0003827.
XX 25-FEB-1997; 97GB-0003850.
XX 28-FEB-1997; 97GB-0004252.
XX
XX (CORT-) CORTECS UK LTD.
XX
XX Engwerda C, Mynott TL, Peek K;
XX
XX WPI; 1998-481194/41.
XX
XX New isolated component of bromelain - used for treating e.g.

PT autoimmune diseases, transplant rejection, allergic reactions,
PT toxic shock, apoptosis, parasite or pathogen infections or cancer
XX
XX Disclosure; Page 24; 56pp; English.
XX
XX The patent discloses a fraction of bromelain, designated CCS, which
CC (i) contains proteins having molecular weights of 15.07 kD, 25.85 kD
CC and 27.45 kD (as determined by SDS-PAGE), (ii) has isoelectric points
CC of 10.4 and 10.45, and (iii) is obtainable by: (a) dissolving bromelain
CC in acetate buffer at pH 5.0; (b) separating the components by fast flow
CC high performance chromatography on S-Sepharose (RPM), eluting with a
CC linear gradient of 0 to 0.8M NaCl in acetate buffer over 300 ml; (c)
CC collecting the fraction corresponding to the final double peak off the
CC column; and (d) isolating the protein from the fraction collected in (c).
CC The CCS fraction can be used: (a) for modulating intracellular signalling
CC pathways which control cell growth and proliferation; (b) for inhibiting
CC the production of growth factors and cytokines by cells; (c) for reducing
CC or preventing the activation of mitogen-activated protein (MAP) kinase
CC pathway; (d) for reducing or preventing the activation of T cells;
CC (e) as an immunosuppressive agent; (f) for blocking the production of
CC growth factors and other cytokines; or (g) for the treatment or
CC prevention of autoimmune diseases, graft or transplant rejection,
CC allergic reactions, toxic shock, apoptosis, parasite or pathogen
CC infections, or cancer. The CCS fraction is postulated to consist of
CC the cysteine proteases ananain and comosain, together with various other
CC components. The present sequence represents the N-terminal of ananain.
XX
XX Sequence 20 AA;
SQ
Query Match 86.1%; Score 93; DB 19; Length 20;
Best Local Similarity 94.7%; Pred. No. 4.4e-09;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 VPOSIDMRDYGAVTSVKNO 19
DB 1 VPOSIDMRDYGAVTSVKNO 19
RESULT 14
AAV51582
ID AAV51582 standard; peptide; 20 AA.
XX
XX AAV51582;
XX
XX 18-MAY-2000 (first entry)
XX
XX Pineapple bromelain protease peptide fragment from fraction 5.
XX
XX Bromelain; pineapple; protease; blood coagulation; inhibitor; plasmin.
XX
XX Ananas comosus.
XX
XX W0200003729-A1.
XX
XX 27-JAN-2000.
XX
XX 15-JUL-1998; 98WO-EP04406.
XX
XX 15-JUL-1998; 98WO-EP04406.
XX
XX (URSA-) URSAFARM ARZNEIMITTEL GMBH.
XX
XX Maurer R, Eckert K, Grabowska E, Bachmann K;
XX
XX WPI; 2000-182338/16.
XX
XX Composition for inhibiting blood coagulation more efficiently than
PT crude bromelain powder comprises bromelain proteinases -
XX
XX Claim 8; Page 16; 24pp; German.
XX
XX This invention describes a novel use of bromelain proteinases (i) for
CC inhibiting blood coagulation. (i) stimulate formation of plasmin, but

CC inhibit both formation of fibrin and adhesion of thrombocytes to
 CC endothelial cells. (1) are about 10 times more active than bromelain
 CC crude powder, i.e. comparable with streptokinase. AAY51581 and AAY51582
 CC represent bromelain proteases isolated from pineapple (Ananas comosus)
 CC which are used in the method of the invention.

XX
 SQ Sequence 20 AA;
 Query Match 86.1%; Score 93; DB 21; Length 20;
 Best Local Similarity 94.7%; Pred. No. 4.4e-09;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 VPOSIDWRDYGAVTSVKNQ 19
 DB 1 VPOSIDWRDYGAVTSVKNQ 19

RESULT 15
 AAY84395
 ID AAY84395 standard; peptide; 21 AA.
 XX
 AC AAY84395;
 XX
 DT 12-JUL-2000 (first entry)
 XX
 DE N-terminal peptide from a 26 kDa protein of bromelain.
 XX
 KW Bromelain; CCS/P2; ERK-2 phosphorylation; interleukin-2;
 KW CD4+ T cell proliferation; mitogen activated protein kinase; MAP kinase;
 KW growth factor; cytokine; cancer; parasite; pathogen infection;
 KW immunosuppressant; autoimmune disease; graft rejection;
 KW transplant rejection; allergic reaction; toxic shock; apoptosis.
 XX
 OS Ananas comosus.

XX
 FH Key Location/Qualifiers
 FT Misc-difference 9 /note= "not specified"
 FT Misc-difference 10 /note= "not specified"
 FT
 PN WO200014254-A2.
 XX
 PD 16-MAR-2000.
 XX
 PF 24-AUG-1999; 99WO-GB02792.
 XX
 PR 02-SEP-1998; 98GB-0019137.
 XX
 PA (CORT-) CORTECS UK LTD.
 XX
 PI Mynott TL, Peek K, Tyreman DR;
 XX
 DR WPI; 2000-256999/22.

XX Bromelain fractions useful in the prevention and treatment of cancer,
 XX parasites and pathogen infections and autoimmune diseases comprise
 XX proteins with biological activity -

PS Claim 5; Page 52; 69pp; English.
 XX
 CC The present sequence represents a N-terminal sequence from a 24 kDa
 CC protein (as measured by SDS-PAGE). The protein is derived from a fraction
 CC of bromelain, designated CCS/P2, comprising proteins of molecular weights
 CC 23.9 kDa and 26.6 kDa when measured by SDS-PAGE or 23.459 kDa and 23.6576
 CC kDa when measured by mass spectrometry. The fraction blocks ERK-2
 CC (serine/threonine kinase) phosphorylation and MAP kinase cascade. They
 CC also block interleukin-2 (IL-2) production and CD4+ T cell proliferation.
 CC The bromelain fractions and proteins are used in medicine as they inhibit
 CC the mitogen activated protein (MAP) kinase pathway, activate T cells and
 CC produce growth factors and cytokines. They can be used for treating
 CC cancer, parasites or pathogen infections. They can be used as an
 CC immunosuppressant for treating or preventing autoimmune diseases, graft

CC or transplant rejection by a host, allergic reactions, toxic shock or
 CC apoptosis. The fractions and proteins may also be used to prepare agents
 CC for the treatment of the above diseases and conditions.

XX
 SQ Sequence 21 AA;
 Query Match 86.1%; Score 93; DB 21; Length 21;
 Best Local Similarity 90.0%; Pred. No. 4.6e-09;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 VPOSIDWRDYGAVTSVKQN 20
 DB 1 VPOSIDWRDYGAVTSVKQN 20

Search completed: November 21, 2002, 16:08:39
 Job time : 32 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2002, 15:59:31 ; Search time 31 Seconds
(without alignments)
85.968 Million cell updates/sec

Title: us-09-674-738-2

Perfect score: 105

Sequence: 1 VPOSIDRDSGAVTSVKNQG 20

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq_101002.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
24: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	105	100.0	20 19 AAW80801	Ananain N-terminal
2	105	100.0	20 21 AAY81582	Pineapple bromelain
3	105	100.0	28 21 AAY84389	N-terminal peptide
4	105	100.0	35 21 AAY84391	N-terminal peptide
5	100	95.2	21 21 AAY84394	N-terminal peptide
6	94	89.5	20 19 AAW80802	Comosain N-terminal
7	93	88.6	20 21 AAY81581	Pineapple bromelain
8	93	88.6	213 15 AAY847550	Stem bromelain pro
9	93	88.6	213 16 AAW86974	Pineapple stem pro
10	88	83.8	21 21 AAY84395	N-terminal peptide

11	85	81.0	21 21 AAY84396	N-terminal peptide
12	84	80.0	311 21 AAY83105	710 proteinase fro
13	84	80.0	324 21 AAY83107	D385848 proteinase
14	84	80.0	351 21 AAY83104	674 proteinase fro
15	83	79.0	377 20 AAW89556	Trifolium sp. cyste
16	82	78.1	127 22 AAW85793	Cysteine protease
17	82	78.1	148 22 AAW85784	Cysteine protease
18	82	78.1	278 22 AAW85782	Cysteine protease
19	81	77.1	258 21 AAG44292	Arabidopsis thalia
20	81	77.1	277 21 AAG30215	Arabidopsis thalia
21	81	77.1	282 21 AAG30214	Arabidopsis thalia
22	81	77.1	294 21 AAG21091	Arabidopsis thalia
23	81	77.1	297 21 AAG41091	Arabidopsis thalia
24	81	77.1	300 21 AAG44290	Arabidopsis thalia
25	81	77.1	361 21 AAG30213	Arabidopsis thalia
26	81	77.1	361 21 AAG50423	Arabidopsis thalia
27	81	77.1	362 21 AAG50422	Arabidopsis thalia
28	81	77.1	368 21 AAG25448	Arabidopsis thalia
29	81	77.1	368 21 AAG42901	Arabidopsis thalia
30	81	77.1	368 22 AAW65770	Cysteine protease
31	80	76.2	129 22 AAW65776	Cysteine protease
32	80	76.2	186 21 AAY83098	569 proteinase fro
33	80	76.2	186 21 AAY83101	573 proteinase fro
34	80	76.2	187 21 AAY83099	571 proteinase fro
35	80	76.2	188 21 AAY83100	572 proteinase fro
36	80	76.2	190 21 AAY83106	D14059 proteinase
37	80	76.2	442 18 AAW27442	Oil seed rape cyst
38	80	76.2	443 18 AAW27441	N-terminal peptide
39	79	75.2	15 21 AAY84397	Bromelain protein
40	79	75.2	18 19 AAW64285	N-terminal peptide
41	79	75.2	20 21 AAY84393	Bromelain protein
42	79	75.2	21 19 AAW64387	N-terminal peptide
43	79	75.2	21 21 AAY84387	Pathogen response
44	79	75.2	231 19 AAW72388	Carica cysteine pr
45	79	75.2	345 22 AAW85128	

ALIGNMENTS

RESULT 1
AAW80801 standard; peptide: 20 AA.
ID AAW80801;
AC AAW80801;
XX
XX
DT 24-NOV-1998 (first entry)
XX
XX
DE Ananain N-terminal sequence.
XX
XX
KW Ananain; bromelain; CCG; autoimmuno; immunosuppressant;
KW graft rejection; antiallergic; antiapoptosis; toxic shock;
KW antiparasitic; antipathogenic; anticancer; antitumor.
XX
XX
OS Ananas comosus.
XX
XX
PN WO9838291-A1.
XX
XX
PD 03-SEP-1998.
XX
XX
PF 25-FEB-1998; 98WO-GB00590.
XX
XX
PR 25-MAR-1997; 97GB-0006119.
PR 25-FEB-1997; 97GB-0003827.
PR 25-FEB-1997; 97GB-0003850.
PR 28-FEB-1997; 97GB-0004252.
XX
XX
PA (CORT-) CORTECS UK LTD.
XX
XX
PI Engwerda C, Mynott TL, Peek K;
XX
XX
DR WPI; 1998-481194/41.
XX

PT New isolated component of bromelain - used for treating e.g.
 PT autoimmune diseases, transplant rejection, allergic reactions,
 PT toxic shock, apoptosis, parasite or pathogen infections or cancer
 XX Disclosure; Page 24; 56pp; English.
 XX The patent discloses a fraction of bromelain, designated CCS, which
 CC (i) contains proteins having molecular weights of 15.07 kD, 25.85 kD
 CC and 27.45 kD (as determined by SDS-PAGE), (ii) has isoelectric points
 CC of 10.4 and 10.45, and (iii) is obtainable by: (a) dissolving bromelain
 CC in acetate buffer at pH 5.0; (b) separating the components by fast flow
 CC high performance chromatography on S-Sepharose (RTM), eluting with a
 CC linear gradient of 0 to 0.6M NaCl in acetate buffer over 300 ml; (c)
 CC collecting the fraction corresponding to the final double peak off the
 CC column; and (d) isolating the protein from the fraction collected in (c).
 CC The CCS fraction can be used: (a) for modulating intracellular signalling
 CC pathways which control cell growth and proliferation; (b) for inhibiting
 CC the production of growth factors and cytokines by cells; (c) for reducing
 CC or preventing the activation of mitogen-activated protein (MAP) kinase
 CC pathway; (d) for reducing or preventing the activation of T cells;
 CC (e) as an immunosuppressive agent; (f) for blocking the production of
 CC growth factors and other cytokines; or (g) for the treatment or
 CC prevention of autoimmune diseases, graft or transplant rejection,
 CC allergic reactions, toxic shock, apoptosis, parasite or pathogen
 CC infections, or cancer. The CCS fraction is postulated to consist of
 CC the cysteine proteases ananain and comosain, together with various other
 CC components. The present sequence represents the N-terminal of ananain.
 XX
 XX Sequence 20 AA;
 SQ Query Match 100.0%; Score 105; DB 19; Length 20;
 Best Local Similarity 100.0%; Pred. No. 1.6e-10;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VPQSIDWRDSCGAVTSVKNOG 20
 DB 1 VPQSIDWRDSCGAVTSVKNOG 20
 RESULT 2
 ID AAY51582 standard; peptide; 20 AA.
 AC AAY51582;
 XX 18-MAY-2000 (first entry)
 XX Pineapple bromelain protease peptide fragment from fraction 5.
 XX Bromelain; pineapple; protease; blood coagulation; inhibitor; plasmin.
 XX Ananas comosus.
 OS WO200003729-A1.
 PN 27-JAN-2000.
 XX 15-JUL-1998; 98WO-EP04406.
 XX 15-JUL-1998; 98WO-EP04406.
 XX (URSA-) URSAPHARM ARZNEIMITTEL GMBH.
 XX Maurer R, Eckert K, Grabowska E, Eschmann K;
 XX WPI; 2000-182338/16.
 XX Composition for inhibiting blood coagulation more efficiently than
 FT crude bromelain powder comprises bromelain proteases -
 XX
 XX Claim 8; Page 16; 24pp; German.
 XX This invention describes a novel use of bromelain proteases (I) for

CC inhibiting blood coagulation. (I) stimulate formation of plasmin, but
 CC inhibit both formation of fibrin and adhesion of thrombocytes to
 CC endothelial cells. (I) are about 10 times more active than bromelain
 CC crude powder, i.e. comparable with streptokinase. AAY51581 and AAY51582
 CC represent bromelain proteases isolated from pineapple (Ananas comosus)
 CC which are used in the method of the invention.
 XX
 XX Sequence 20 AA;
 SQ Query Match 100.0%; Score 105; DB 21; Length 20;
 Best Local Similarity 100.0%; Pred. No. 1.6e-10;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VPQSIDWRDSCGAVTSVKNOG 20
 DB 1 VPQSIDWRDSCGAVTSVKNOG 20
 RESULT 3
 ID AAY84389 standard; peptide; 28 AA.
 XX AAY84389;
 AC AAY84389;
 XX 12-JUL-2000 (first entry)
 XX N-terminal peptide from a fraction of bromelain designated CCS/P2.
 DE
 XX Bromelain; CCS/P2; ERK-2 phosphorylation; interleukin-2;
 XX CD4+ T cell proliferation; mitogen activated protein kinase; MAP kinase;
 XX growth factor; cytokine; cancer; parasite; pathogen infection;
 XX immunosuppressant; autoimmune disease; graft rejection;
 XX transplant rejection; allergic reaction; toxic shock; apoptosis.
 XX Ananas comosus.
 OS
 XX Key Location/Qualifiers
 FH Misc-difference 22 note= "not specified"
 FT Misc-difference 25 /note= "not specified"
 FT Misc-difference 26 /note= "not specified"
 FT Misc-difference 26 /note= "not specified"
 XX WO200014254-A2.
 PN 16-MAR-2000.
 XX 24-AUG-1999; 99WO-GB02792.
 XX 02-SEP-1998; 98GB-0019137.
 XX (CORT-) CORTECS UK LTD.
 XX Mynott TL, Peek K, Tyreman DR;
 XX WPI; 2000-256999/22.
 XX Bromelain fractions useful in the prevention and treatment of cancer,
 PT parasites and pathogen infections and autoimmune diseases comprise
 PT proteins with biological activity -
 XX Claim 1; Page 51; 69pp; English.
 XX The present sequence represents a N-terminal sequence from a fraction
 CC of bromelain, designated CCS/P2, comprising proteins of molecular weights
 CC 23.9 kDa and 26.6 kDa when measured by SDS-PAGE or 23.459 kDa and 23.6576
 CC kDa when measured by mass spectrometry. The peptides block ERK-2
 CC (serine/threonine kinase) phosphorylation and MAP kinase cascade. They
 CC also block interleukin-2 (IL-2) production and CD4+ T cell proliferation.
 CC The bromelain fractions and proteins are used in medicine as they inhibit
 CC the mitogen activated protein (MAP) kinase pathway, activate T cells and
 CC produce growth factors and cytokines. They can be used for treating

CC cancer, parasites or pathogen infections. They can be used as an
 CC immunosuppressant for treating or preventing autoimmune diseases, graft
 CC or transplant rejection by a host, allergic reactions, toxic shock or
 CC apoptosis. The fractions and proteins may also be used to prepare agents
 CC for the treatment of the above diseases and conditions.

XX Sequence 28 AA;

Query Match 100.0%; Score 105; DB 21; Length 28;

Best Local Similarity 100.0%; Pred. No. 2.3e-10; Mismatches 0; Gaps 0;

Matches 20; Conservative 0; Indels 0; Gaps 0;
 Oy 1 VPOSIDMRDSCGAVTSVKNQ 20
 1 VPOSIDMRDSCGAVTSVKNQ 20
 Db 1 VPOSIDMRDSCGAVTSVKNQ 20

RESULT 4

ID AAY84391 standard; peptide; 35 AA.

XX AAY84391;

XX 12-JUL-2000 (first entry)

DE N-terminal peptide from a fraction of bromelain designated CCS/P2.

XX Bromelain; CCS/P2; ERK-2 phosphorylation; interleukin-2;

XX CD4+ T cell proliferation; mitogen activated protein kinase; MAP kinase;

XX growth factor; cytokine; cancer; parasite; pathogen infection;

XX immunosuppressant; autoimmune disease; graft rejection;

XX transplant rejection; allergic reaction; toxic shock; apoptosis.

XX Ananas comosus.

XX Key Location/Qualifiers

FT Misc-difference 25 /note= "not specified"

FT Misc-difference 29 /note= "not specified"

FT Misc-difference 32 /note= "not specified"

FT Misc-difference 33 /note= "not specified"

FT Misc-difference 34 /note= "not specified"

XX WO200014254-A2.

XX 16-MAR-2000.

XX 24-AUG-1999; 99WO-GB02792.

XX 02-SEP-1998; 98GB-0019137.

XX (CORT-) CORTECS UK LTD.

XX Mynott TL, Peek K, Tyreman DR;

XX WPI; 2000-256999/22.

XX Bromelain fractions useful in the prevention and treatment of cancer,

XX parasites and pathogen infections and autoimmune diseases comprise

XX proteins with biological activity -

XX Claim 1; Page 51; 69pp; English.

XX The present sequence represents a N-terminal sequence from a fraction

XX of bromelain, designated CCS/P2, comprising proteins of molecular weights

XX 23.9 kDa and 26.6 kDa when measured by SDS-PAGE or 23.459 kDa and 23.6576

XX kDa when measured by mass spectrometry. The peptides block ERK-2

XX (serine/threonine kinase) phosphorylation and MAP kinase cascade. They

XX also block interleukin-2 (IL-2) production and CD4+ T cell proliferation.

CC The bromelain fractions and proteins are used in medicine as they inhibit
 CC the mitogen activated protein (MAP) kinase pathway, activate T cells and
 CC produce growth factors and cytokines. They can be used for treating
 CC cancer, parasites or pathogen infections. They can be used as an
 CC immunosuppressant for treating or preventing autoimmune diseases, graft
 CC or transplant rejection by a host, allergic reactions, toxic shock or
 CC apoptosis. The fractions and proteins may also be used to prepare agents
 CC for the treatment of the above diseases and conditions.

XX Sequence 35 AA;

Query Match 100.0%; Score 105; DB 21; Length 35;

Best Local Similarity 100.0%; Pred. No. 3e-10; Mismatches 0; Gaps 0;

Matches 20; Conservative 0; Indels 0; Gaps 0;
 Oy 1 VPOSIDMRDSCGAVTSVKNQ 20
 1 VPOSIDMRDSCGAVTSVKNQ 20
 Db 1 VPOSIDMRDSCGAVTSVKNQ 20

RESULT 5

ID AAY84394 standard; peptide; 21 AA.

XX AAY84394;

XX 12-JUL-2000 (first entry)

DE N-terminal peptide from a 24 kDa protein of bromelain.

XX Bromelain; CCS/P2; ERK-2 phosphorylation; interleukin-2;

XX CD4+ T cell proliferation; mitogen activated protein kinase; MAP kinase;

XX growth factor; cytokine; cancer; parasite; pathogen infection;

XX immunosuppressant; autoimmune disease; graft rejection;

XX transplant rejection; allergic reaction; toxic shock; apoptosis.

XX Ananas comosus.

XX Key Location/Qualifiers

FT Misc-difference 16 /note= "not specified"

XX WO200014254-A2.

XX 16-MAR-2000.

XX 24-AUG-1999; 99WO-GB02792.

XX 02-SEP-1998; 98GB-0019137.

XX (CORT-) CORTECS UK LTD.

XX Mynott TL, Peek K, Tyreman DR;

XX WPI; 2000-256999/22.

XX Bromelain fractions useful in the prevention and treatment of cancer,

XX parasites and pathogen infections and autoimmune diseases comprise

XX proteins with biological activity -

XX Claim 3; Page 52; 69pp; English.

XX The present sequence represents a N-terminal sequence from a 24 kDa

XX protein (as measured by SDS-PAGE). The protein is derived from a fraction

XX of bromelain, designated CCS/P2, comprising proteins of molecular weights

XX 23.9 kDa and 26.6 kDa when measured by SDS-PAGE or 23.459 kDa and 23.6576

XX kDa when measured by mass spectrometry. The fraction blocks ERK-2

XX (serine/threonine kinase) phosphorylation and MAP kinase cascade. They

XX also block interleukin-2 (IL-2) production and CD4+ T cell proliferation.

XX The bromelain fractions and proteins are used in medicine as they inhibit

XX the mitogen activated protein (MAP) kinase pathway, activate T cells and

XX produce growth factors and cytokines. They can be used for treating

XX cancer, parasites or pathogen infections. They can be used as an

CC immunosuppressant for treating or preventing autoimmune diseases, graft
 CC or transplant rejection by a host, allergic reactions, toxic shock or
 CC apoptosis. The fractions and proteins may also be used to prepare agents
 CC for the treatment of the above diseases and conditions.

XX
 XX
 SQ Sequence 21 AA;

Query Match 95.2%; Score 100; DB 21; Length 21;
 Best Local Similarity 95.0%; Pred. No. 1.1e-09;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPOSIDWRDVGAVTSVKNQ 20
 |||||
 DB 1 VPOSIDWRDVGAVTSVKNQ 20

RESULT 6
 AAW80802
 ID AAW80802 standard; peptide; 20 AA.
 XX
 AC AAW80802;
 XX
 DT 24-NOV-1998 (first entry)
 XX
 DE Comosain N-terminal sequence.
 XX
 KW Comosain; bromelain; CCS; autoimmune; immunosuppressant;
 KW graft rejection; anti-allergic; antiapoptosis; toxic shock;
 KW antiparasitic; antipathogenic; anticancer; antitumour.
 XX
 OS Ananas comosus.
 XX
 PN WO9838291-Al.
 XX
 PD 03-SEP-1998.
 XX
 PF 25-FEB-1998; 98WO-GB00590.
 XX
 PR 25-MAR-1997; 97GB-0006119.
 PR 25-FEB-1997; 97GB-0003827.
 PR 25-FEB-1997; 97GB-0003850.
 PR 28-FEB-1997; 97GB-0004252.
 XX
 XX (CORT-) CORTECS UK LTD.
 XX
 XX Engwerda C, Mynott TL, Peek K;
 XX
 XX WPI; 1998-481194/41.
 XX
 XX New isolated component of bromelain - used for treating e.g.
 PT autoimmune diseases, transplant rejection, allergic reactions,
 PT toxic shock, apoptosis, parasite or pathogen infections or cancer
 XX
 XX Disclosure: Page 24; 56pp; English.

The patent discloses a fraction of bromelain, designated CCS, which
 CC (i) contains proteins having molecular weights of 15.07 kD, 25.85 kD
 CC and 27.45 kD (as determined by SDS-PAGE), (ii) has isoelectric points
 CC of 10.4 and 10.45, and (iii) is obtainable by: (a) dissolving bromelain
 CC in acetate buffer at pH 5.0; (b) separating the components by fast flow
 CC high performance chromatography on S-Sepharose (RPM), eluting with a
 CC linear gradient of 0 to 0.8M NaCl in acetate buffer over 300 ml; (c)
 CC collecting the fraction corresponding to the final double peak off the
 CC column; and (d) isolating the protein from the fraction collected in (c).
 CC The CCS fraction can be used: (a) for modulating intracellular signalling
 CC pathways which control cell growth and proliferation; (b) for inhibiting
 CC the production of growth factors and cytokines by cells; (c) for reducing
 CC or preventing the activation of mitogen-activated protein (MAP) kinase
 CC pathway; (d) for reducing or preventing the activation of T cells;
 CC (e) as an immunosuppressive agent; (f) for blocking the production of
 CC growth factors and other cytokines; or (g) for the treatment or
 CC prevention of autoimmune diseases, graft or transplant rejection,
 CC allergic reactions, toxic shock, apoptosis, parasite or pathogen

CC infections, or cancer. The CCS fraction is postulated to consist of
 CC the cysteine proteases ananain and comosain, together with various other
 CC components. The present sequence represents the N-terminal of comosain.

XX
 XX
 SQ Sequence 20 AA;

Query Match 89.5%; Score 94; DB 19; Length 20;
 Best Local Similarity 90.0%; Pred. No. 1.1e-08;
 Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPOSIDWRDVGAVTSVKNQ 20
 |||||
 DB 1 VPOSIDWRDVGAVTSVKNQ 20

RESULT 7
 AAY51581
 ID AAY51581 standard; peptide; 20 AA.
 XX
 AC AAY51581;
 XX
 DT 18-MAY-2000 (first entry)
 XX
 DE Pineapple bromelain protease peptide fragment from fraction 4.
 XX
 KW Bromelain; pineapple; protease; blood coagulation; inhibitor; plasmin.
 XX
 OS Ananas comosus.
 XX
 PN WO2000003729-Al.
 XX
 XX WO2000003729-Al.
 XX
 PD 27-JAN-2000.
 XX
 PF 15-JUL-1998; 98WO-EP04406.
 XX
 PR 15-JUL-1998; 98WO-EP04406.
 XX
 XX (URSA-) URSAPHARM ARZNEIMITTEL GMBH.
 PA
 PI Maurer R, Eckert K, Grabowska E, Eschmann K;
 XX
 DR WPI; 2000-182338/16.
 XX
 XX Composition for inhibiting blood coagulation more efficiently than
 PT crude bromelain powder comprises bromelain proteases -
 XX
 XX Claim 6; Page 15; 24pp; German.

This invention describes a novel use of bromelain proteases (I) for
 CC inhibiting blood coagulation. (I) stimulate formation of plasmin, but
 CC inhibit both formation of fibrin and adhesion of thrombocytes to
 CC endothelial cells. (I) are about 10 times more active than bromelain
 CC crude powder, i.e. comparable with streptokinase. AAY51581 and AAY51582
 CC represent bromelain proteases isolated from pineapple (Ananas comosus)
 CC which are used in the method of the invention.

XX
 XX
 SQ Sequence 20 AA;

Query Match 88.6%; Score 93; DB 21; Length 20;
 Best Local Similarity 94.7%; Pred. No. 1.6e-08;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPOSIDWRDVGAVTSVKNQ 19
 |||||
 DB 1 VPOSIDWRDVGAVTSVKNQ 19

RESULT 8
 AAR47550
 ID AAR47550 standard; protein; 213 AA.
 XX
 AC AAR47550;
 XX

DT 18-JUL-1994 (first entry)
 XX Stem bromelain protease.
 DE
 XX Proteolytic enzyme; cysteine protease; diarrhoea; fluid secretion;
 KM gastro-intestinal tract; GI tract; absorption.
 XX
 OS Ananas comosus.
 XX
 PN WO9400147-A.
 XX
 PD 06-JAN-1994.
 XX
 PF 30-JUN-1993; 93WO-GB01374.
 XX
 PR 30-JUN-1992; 92GB-0013862.
 PR 20-APR-1993; 93GB-0008164.
 PR 25-JUN-1993; 93GB-0013189.
 XX
 PA (CORT-) CORTECS LTD.
 XX
 PI Mynott TL;
 XX
 DR WPI; 1994-025889/03.
 XX
 PT Enzyme, esp. bromelain, contg. compsn. for control or diarrhoea -
 PT esp. when caused by heat-labile toxin-producing E. coli, also
 PT inhibits fluid secretion in the gut and promotes absorption
 XX
 PS Disclosure; Page 11-13; 70pp; English.
 XX
 CC The sequence is that of stem bromelain protease, a mixture of
 CC various moieties derived from the stem of the pineapple. The
 CC protein is a cysteine protease used to treat infectious diarrhoea
 CC caused by heat labile or heat stable toxins, enterotoxigenic E. coli
 CC and Vibrio cholerae. The enzyme reduces toxin binding activity,
 CC inhibits the secretory effect of the toxins and prevents attachment
 CC of bacteria to the intestinal lining. It also inhibits secretion
 CC caused by secretagogues, e.g. PGE2, theophylline, etc. and has no
 CC adverse effect on nutrient influx.
 CC
 SQ Sequence 213 AA;
 QY
 Db 1 VPOSIDMRDYGAVTSYKNO 19
 2 VPOSIDMRDYGAVTSYKNO 20
 RESULT 9
 AAR68974
 ID AAR68974 standard; protein; 213 AA.
 XX
 AC AAR68974;
 XX
 DT 09-AUG-1995 (first entry)
 XX
 DE Pineapple stem bromelain protease.
 XX
 KM Bromelain; protease; proteolytic enzyme; cholera; diphtheria; cancer;
 KM inflammation.
 XX
 OS Ananas comosus.
 XX
 PN WO9500163-A.
 XX
 PD 05-JAN-1995.
 XX
 PF 24-JUN-1994; 94WO-GB01368.
 XX

PR 25-JUN-1993; 93GB-0013188.
 XX
 PA (CORT-) CORTECS LTD.
 XX
 PI Mynott TL;
 XX
 DR WPI; 1995-051751/07.
 XX
 PT Use of purified stem bromelain protease - to treat conditions
 PT dependent on cyclic nucleotide pathways e.g. cholera, diphtheria
 PT or cancer
 XX
 PS Claim 1; Page 35; 45pp; English.
 XX
 CC Pineapple stem bromelain protease is used in the treatment or
 CC prophylaxis of conditions which are dependent on cyclic nucleotide
 CC and other intracellular and/or extracellular pathways. The diseases
 CC include cancer, inflammation, atherosclerosis and bacterial
 CC infections such as cholera, diphtheria and pertussis.
 CC
 SQ Sequence 213 AA;
 QY
 Db 1 VPOSIDMRDYGAVTSYKNO 19
 2 VPOSIDMRDYGAVTSYKNO 20
 RESULT 10
 AAY84395
 ID AAY84395 standard; peptide; 21 AA.
 XX
 AC AAY84395;
 XX
 DT 12-JUL-2000 (first entry)
 XX
 DE N-terminal peptide from a 26 kDa protein of bromelain.
 XX
 KM Bromelain; CCS/P2; ERK-2 phosphorylation; interleukin-2;
 KM CD4+ T cell proliferation; mitogen activated protein kinase; MAP kinase;
 KM growth factor; cytokine; cancer; parasite; pathogen infection;
 KM immunosuppressant; autoimmune disease; graft rejection;
 KM transplant rejection; allergic reaction; toxic shock; apoptosis.
 XX
 OS Ananas comosus.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 9 /note= "not specified"
 FT Misc-difference 10 /note= "not specified"
 FT
 XX
 PN WO200014254-A2.
 XX
 PD 16-MAR-2000.
 XX
 PF 24-AUG-1999; 99WO-GB02792.
 XX
 PR 02-SEP-1998; 98GB-0019137.
 XX
 PA (CORT-) CORTECS UK LTD.
 XX
 PI Mynott TL, Peek K, Tyreman DR;
 XX
 DR WPI; 2000-256999/22.
 XX
 PT Bromelain fractions useful in the prevention and treatment of cancer,
 PT parasites and pathogen infections and autoimmune diseases comprise
 PT proteins with biological activity -
 XX

PS Claim 5; Page 52; 69pp; English.

XX The present sequence represents a N-terminal sequence from a 24 kDa protein (as measured by SDS-PAGE). The protein is derived from a fraction of bromelain, designated CCS/P2, comprising proteins of molecular weights 23.9 kDa and 26.6 kDa when measured by SDS-PAGE or 23.459 kDa and 23.6576 kDa when measured by mass spectrometry. The fraction blocks ERK-2 (serine/threonine kinase) phosphorylation and MAP kinase cascade. They also block interleukin-2 (IL-2) production and CD4+ T cell proliferation. The bromelain fractions and proteins are used in medicine as they inhibit the mitogen activated protein (MAP) kinase pathway, activate T cells and produce growth factors and cytokines. They can be used for treating cancer, parasites or pathogen infections. They can be used as an immunosuppressant for treating or preventing autoimmune diseases, graft or transplant rejection by a host, allergic reactions, toxic shock or apoptosis. The fractions and proteins may also be used to prepare agents for the treatment of the above diseases and conditions.

SQ Sequence 21 AA;

Query Match 83.8%; Score 88; DB 21; Length 21;

Best Local Similarity 89.5%; Pred. No. 1.1e-07;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPQSIDWRDVGAVTSVKNQ 19
||||||| |||||||

Db 1 VPQSIDWRDVGAVTSVKNQ 19

RESULT 11

AA484396

ID AA484396 standard; peptide; 21 AA.

AC AA484396;

DT 12-JUL-2000 (first entry)

DE N-terminal peptide from a 24 kDa protein of bromelain.

XX Bromelain; CCS/P2; ERK-2 phosphorylation; interleukin-2;

KW CD4+ T cell proliferation; mitogen activated protein kinase; MAP kinase;

KW growth factor; cytokine; cancer; parasite; pathogen infection;

KW immunosuppressant; autoimmune disease; graft rejection;

KW transplant rejection; allergic reaction; toxic shock; apoptosis.

XX Ananas comosus.

OS Ananas comosus.

XX Key Location/Qualifiers

FH Misc-difference 17 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

FT Misc-difference 19 /note= "not specified"

CC The present sequence represents a N-terminal sequence from a 24 kDa protein (as measured by SDS-PAGE). The protein is derived from a fraction of bromelain, designated CCS/P2, comprising proteins of molecular weights 23.9 kDa and 26.6 kDa when measured by SDS-PAGE or 23.459 kDa and 23.6576 kDa when measured by mass spectrometry. The fraction blocks ERK-2 (serine/threonine kinase) phosphorylation and MAP kinase cascade. They also block interleukin-2 (IL-2) production and CD4+ T cell proliferation. The bromelain fractions and proteins are used in medicine as they inhibit the mitogen activated protein (MAP) kinase pathway, activate T cells and produce growth factors and cytokines. They can be used for treating cancer, parasites or pathogen infections. They can be used as an immunosuppressant for treating or preventing autoimmune diseases, graft or transplant rejection by a host, allergic reactions, toxic shock or apoptosis. The fractions and proteins may also be used to prepare agents for the treatment of the above diseases and conditions.

SQ Sequence 21 AA;

Query Match 81.0%; Score 85; DB 21; Length 21;

Best Local Similarity 85.0%; Pred. No. 3.5e-07;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPQSIDWRDVGAVTSVKNQ 20

Db 1 VPQSIDWRDVGAVTSVKPXG 20

RESULT 12

AA483105

ID AA483105 standard; Protein; 311 AA.

AC AA483105;

DT 04-JUL-2000 (first entry)

DE 710 proteinase from pineapple stem.

XX Proteinase; pineapple; treatment; prophylaxis; tumour;

XX solid tumour; non-solid tumour; virus induced tumour;

XX immunosuppressant; immunodeficiency; HIV;

XX human immunodeficiency virus; malacia; malnutrition; trauma; burn;

XX surgery; genetic abnormality; diabetes; nitric oxide; NO;

XX antimicrobial.

XX Ananas comosus.

XX WO200014253-A1.

XX 16-MAR-2000.

XX 24-AUG-1999; 99WO-GB02784.

XX 02-SEP-1998; 98GB-0019138.

XX (CORT-) CORTECS UK LTD.

XX Mynott TL, Crossett B;

XX WPI; 2000-256998/22.

XX N-PSDB; AA293389.

XX Stem bromelain-like proteinases useful for the prevention and treatment of cancer and diseases or conditions which respond to increased nitric oxide production are encoded by nucleic acid sequences isolated from a pineapple stem -

XX Claim 4; Page 58-59; 72pp; English.

XX The sequences encoding proteinases from pineapple stem have the same activity as a stem bromelain-like proteinase, CCX2.

XX The proteinases can be used in human or veterinary medicine and can be used for treating or preventing cancer, e.g. solid tumours;

XX non-solid tumours and virus induced tumours. The proteinases can

CC also be used as an immunostimulant or in the preparation of an
CC immunostimulant to treat immunodeficiencies arising from
CC malnutrition, infection (e.g. HIV and malaria), tumours (e.g.
CC lymphoid, myeloma), trauma (e.g. burns, wounds, surgery) medical
CC treatment (e.g. with drugs such as steroids, cyclosporin and
CC cyclophosphamide), protein loss (e.g. in diarrhoea and burns),
CC genetic abnormalities (e.g. in combined immunodeficiency patients
CC lacking B or T cells), diabetes and old age. The proteinase can
CC also be used in the treatment or prevention of diseases or
CC conditions which respond to increased nitric oxide (NO) production,
CC or as an antimicrobial agent or the component of an antimicrobial
CC agent.

XX Sequence 311 AA;

Query Match 80.0%; Score 84; DB 21; Length 311;
Best Local Similarity 84.2%; Pred. No. 1.2e-05;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPOSIDMRDSCGAVTSVKNO 19
DB 96 VPOSIDMRDYGAVNEVKNO 114

RESULT 13

AAV83107 standard; Protein: 324 AA.

XX AAV83107;

DE 04-JUL-2000 (first entry)

XX D385848 proteinase from pineapple stem.

KW Proteinase; pineapple; treatment; prevention; prophylaxis; tumour;
KW solid tumour; non-solid tumour; virus induced tumour;
KW immunostimulant; immunodeficiency; HIV;

KW human immunodeficiency virus; malaria; malnutrition; trauma; burn;
KW surgery; genetic abnormality; diabetes; nitric oxide; NO;
KW antimicrobial.

XX Ananas comosus.

XX WO200014253-A1.

XX 16-MAR-2000.

XX 24-AUG-1999; 99WO-GB02784.

XX 02-SEP-1998; 98GB-0019138.

XX (CORT-) CORTECS UK LTD.

XX Mynott TL, Crosette B;

XX WPI; 2000-256998/22.

XX N-PSDB; AA293391.

PT Stem bromelain-like proteinases useful for the prevention and treatment
PT of cancer and diseases or conditions which respond to increased nitric
PT oxide production are encoded by nucleic acid sequences isolated from a
PT pineapple stem -

XX Claim 4; Page 64-66; 72pp; English.

CC The sequences encoding proteinases from pineapple stem have
CC the same activity as a stem bromelain-like proteinase, CCX2.
CC The proteinases can be used in human or veterinary medicine and
CC can be used for treating or preventing cancer, e.g. solid tumours;
CC non-solid tumours and virus induced tumours. The proteinases can
CC also be used as an immunostimulant or in the preparation of an
CC immunostimulant to treat immunodeficiencies arising from
CC malnutrition, infection (e.g. HIV and malaria), tumours (e.g.

CC lymphoid, myeloma), trauma (e.g. burns, wounds, surgery) medical
CC treatment (e.g. with drugs such as steroids, cyclosporin and
CC cyclophosphamide), protein loss (e.g. in diarrhoea and burns),
CC genetic abnormalities (e.g. in combined immunodeficiency patients
CC lacking B or T cells), diabetes and old age. The proteinase can
CC also be used in the treatment or prevention of diseases or
CC conditions which respond to increased nitric oxide (NO) production,
CC or as an antimicrobial agent or the component of an antimicrobial
CC agent.

XX Sequence 324 AA;

Query Match 80.0%; Score 84; DB 21; Length 324;
Best Local Similarity 84.2%; Pred. No. 1.2e-05;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPOSIDMRDSCGAVTSVKNO 19
DB 96 VPOSIDMRDYGAVNEVKNO 114

RESULT 14

AAV83104 standard; Protein: 351 AA.

XX AAV83104;

DE 04-JUL-2000 (first entry)

XX 674 proteinase from pineapple stem.

KW Proteinase; pineapple; treatment; prevention; prophylaxis; tumour;
KW solid tumour; non-solid tumour; virus induced tumour;
KW immunostimulant; immunodeficiency; HIV;

KW human immunodeficiency virus; malaria; malnutrition; trauma; burn;
KW surgery; genetic abnormality; diabetes; nitric oxide; NO;
KW antimicrobial.

XX Ananas comosus.

XX WO200014253-A1.

XX 16-MAR-2000.

XX 24-AUG-1999; 99WO-GB02784.

XX 02-SEP-1998; 98GB-0019138.

XX (CORT-) CORTECS UK LTD.

XX Mynott TL, Crosette B;

XX WPI; 2000-256998/22.

XX N-PSDB; AA293398.

PT Stem bromelain-like proteinases useful for the prevention and treatment
PT of cancer and diseases or conditions which respond to increased nitric
PT oxide production are encoded by nucleic acid sequences isolated from a
PT pineapple stem -

XX Claim 4; Page 55-57; 72pp; English.

CC The sequences encoding proteinases from pineapple stem have
CC the same activity as a stem bromelain-like proteinase, CCX2.
CC The proteinases can be used in human or veterinary medicine and
CC can be used for treating or preventing cancer, e.g. solid tumours;
CC non-solid tumours and virus induced tumours. The proteinases can
CC also be used as an immunostimulant or in the preparation of an
CC immunostimulant to treat immunodeficiencies arising from
CC malnutrition, infection (e.g. HIV and malaria), tumours (e.g.
CC lymphoid, myeloma), trauma (e.g. burns, wounds, surgery) medical
CC treatment (e.g. with drugs such as steroids, cyclosporin and
CC cyclophosphamide), protein loss (e.g. in diarrhoea and burns),

CC genetic abnormalities (e.g. in combined immunodeficiency patients
 CC lacking B or T cells), diabetes and old age. The proteinase can
 CC also be used in the treatment or prevention of diseases or
 CC conditions which respond to increased nitric oxide (NO) production,
 CC or as an antimicrobial agent or the component of an antimicrobial
 CC agent.
 XX
 SQ Sequence 351 AA;

Query Match 80.0%; Score 84; DB 21; Length 351;
 Best Local Similarity 84.2%; Pred. No. 1.3e-05;
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSDGAVTSVKNG 19
 |||||
 DB 123 VPOSIDWRDYGAVNEVKNG 141

RESULT 15

AAW89556
 ID AAW89556 standard; Protein; 377 AA.

AC AAW89556;

DT 16-MAR-1999 (first entry)

XX Triticum sp. cysteine proteinase #1.

XX Triticum; wheat seed; cysteine proteinase; gluten; baking.

OS Triticum sp.

XX JP10327886-A.

XX 15-DEC-1998.

XX 27-MAR-1998; 98JP-0098140.

XX 31-MAR-1997; 97JP-0114946.

XX (SHOS) SHOWA SANGYO CO.

XX WPI; 1999-109255/10.

DR N-PSDB; AAW82456.

XX New DNA coding cysteine proteinase originating from wheat seed -
 PT useful for improving gluten for use in bakery process

PS Claim 1; Page 15-16; 29pp; Japanese.

XX The present sequence represents a cysteine proteinase isolated from
 CC wheat seed (Triticum sp.). The cysteine proteinase is useful for
 CC improving gluten for use in the bakery process.

XX Sequence 377 AA;

Query Match 79.0%; Score 83; DB 20; Length 377;
 Best Local Similarity 70.0%; Pred. No. 2.1e-05;
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSDGAVTSVKNG 20
 |||||
 DB 138 LPRSDWRQKQKAVTSVKNG 157

Search completed: November 21, 2002, 16:08:39
 Job time : 31 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2002, 15:59:48 ; Search time 7.5 Seconds
(without alignments)
110.604 Million cell updates/sec

Title: US-09-674-738-2
Perfect score: 105
Sequence: 1 VPSIDWRDGAVTGVKNG 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	105	100.0	345	1 ANAN ANACO	P80884 ananas como
2	93	88.6	212	1 BROW ANACO	P14518 ananas como
3	84	80.0	352	1 PAP2 CARPA	P14080 carica papa
4	83	79.0	334	1 CATL_PIG	Q28944 sus scrofa
5	82	78.1	43	1 CC3 CARCN	P32956 carica cand
6	82	78.1	43	1 CC4 CARCN	P32957 carica cand
7	82	78.1	334	1 CATL_MOUSE	P06797 mus musculu
8	81	77.1	319	1 CATL_MOUSE	Q25534 schistosoma
9	81	77.1	361	1 A494 ARATH	P43295 arabidopsis
10	81	77.1	368	1 RD19 ARATH	P43296 arabidopsis
11	80	76.2	221	1 GPDI ZINOF	P82474 zingiber of
12	80	76.2	471	1 ORYB_ORYSA	P25777 oryza sativ
13	79	75.2	221	1 GPI ZINOF	P82473 zingiber of
14	79	75.2	334	1 CATL_RAT	P07154 rattus norv
15	79	75.2	345	1 PAP4 CARPA	P08784 carica papa
16	79	75.2	354	1 CYS1_LEIFI	P35591 leishmania
17	79	75.2	354	1 LCPA_LEIME	P25775 leishmania
18	78	74.3	341	1 CATL_DROME	Q95029 drosophila
19	78	74.3	362	1 CYSF_PHAVU	P25803 phaseolus v
20	78	74.3	362	1 CYSF_VIGMO	P12412 vigna mungo
21	78	74.3	443	1 LCPB_LEIME	P34400 leishmania
22	78	74.3	444	1 CYS2_LEIFI	Q05094 leishmania
23	77	73.3	333	1 CATL_HUMAN	P07711 homo sapien
24	76	72.4	348	1 PAP4 CARPA	P05994 carica papa
25	76	72.4	371	1 CYS1_HORVU	P25240 hordeum vul
26	76	72.4	373	1 CYS2_HORVU	P25250 hordeum vul
27	74	70.5	343	1 CYS1_DICDI	P04988 dictyosteli
28	74	70.5	462	1 RD21 ARATH	P43297 arabidopsis
29	73	69.5	330	1 CATS_RAT	Q02765 rattus norv
30	73	69.5	363	1 CYSF_PPA	P25804 pisum sativ
31	73	69.5	395	1 CATL_BRUPA	O14723 brugia paha
32	72	68.6	217	1 CATL_SHEEP	O10991 ovis aries
33	72	68.6	218	1 CATL_CHICK	P09648 gallus gall

34	72	68.6	331	1 CATS_HUMAN	P25774 homo sapien
35	72	68.6	333	1 TEST_RAT	P15242 rattus norv
36	72	68.6	334	1 CATL_BOVIN	P25975 bos taurus
37	72	68.6	371	1 CYS1_MAIZE	Q10716 zea mays (m
38	72	68.6	458	1 ORYB_ORYSA	P25776 oryza sativ
39	71	67.6	43	1 CCI CARCN	P32954 carica cand
40	71	67.6	348	1 PAP3 CARPA	P10056 carica papa
41	70	66.7	217	1 CATS_BOVIN	P25326 bos taurus
42	70	66.7	328	1 CYS4_BRUPA	P25251 brassica na
43	70	66.7	346	1 CYS1_LYCES	P20721 lycopersico
44	70	66.7	467	1 CYSF_TRYCK	P25779 trypanosoma
45	69	65.7	329	1 CATL_MOUSE	P55097 mus musculu

ALIGNMENTS

RESULT 1	ANAN ANACO	STANDARD;	PRT;	345 AA.
ID	ANAN ANACO			
AC	P80884; 022293;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Ananain precursor (EC 3.4.22.31).			
GN	AN1.			
OS	Ananas comosus (Pineapple).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae; Ananas.			
OC	Bromeliaceae; Ananas.			
OX	NCBI_TaxID=4615;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Smooth Cayenne; TISSUE=Stem;			
RA	Robertson C.E., Goodenough P.W.;			
RT	"Cloning and expression of ananain gene from pineapple";			
RL	Submitted (NOV-1997) to the EMBL/Genbank/DBO databases.			
RN	[2]			
RP	SEQUENCE OF 123-338.			
RC	TISSUE=Stem;			
RX	MEDLINE=97454461; PubMed=9355753;			
RA	Lee K.L., Albee K.L., Bernasconi R.J., Edmunds T.;			
RT	"Complete amino acid sequence of ananain and a comparison with stem bromelain and other plant cysteine proteases.";			
RL	Biochem. J. 327:199-202(1997).			
CC	- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity for peptide bonds. Best reported small molecule substrate Bz-Phe-Val-Arg-NHMe, but broader specificity than fruit bromelain.			
CC	- ENZYME REGULATION: STRONGLY INHIBITED BY CHICKEN EGG-WHITE CYSSTATIN AND THE ACTIVE-SITE-DIRECTED INHIBITOR TRANS-EXOXYSUCCINYL-L-LEUCYLAMIDO-(4-GUANIDINO)BUTANE (E-64).			
CC	- TISSUE SPECIFICITY: STEM.			
CC	- MASS SPECTROMETRY: MW=23478; METHOD=Electrospray.			
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; AJ002477; CAA05487.1; -			
DR	HSSP; P14080; 1VAL.			
DR	MEROPS; C01.026; -			
DR	InterPro; IPR000668; Peptidase_C1.			
DR	InterPro; IPR000169; SHprot_acidite.			
DR	Pfam; PF00112; Peptidase_C1; 1.			
DR	PRINTS; PRO0705; PAPA1N.			
DR	ProDom; PD000158; Peptidase_C1; 1.			
DR	ProSITE; PS00139; THIOL_PROTEASE_CYS; 1.			
DR	ProSITE; PS00639; THIOL_PROTEASE_HIS; 1.			

PROSITE; PS00640; THIOI_PROTEASE_ASN; 1.
Hydrolase; thiol protease; zymogen; Signal.
FT SIGNAL 1 24
FT PROPEP 25 122
FT CHAIN 123 345
FT ACT SITE 127 147
FT ACT SITE 149 179
FT ACT SITE 179 279
FT DISULFID 144 184
FT DISULFID 178 217
FT DISULFID 273 325
FT CONFLICT 291 291
FT CONFLICT 324 324
FT SEQUENCE 345 AA; 38248 MW; FAF2989080174D87 CRC64;

Query Match 100.0%; Score 105; DB 1; Length 345;
Best Local Similarity 100.0%; Pred. No. 4.4e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSCGAVTSVKNQ 20
DB 123 VPOSIDWRDSCGAVTSVKNQ 142

RESULT 2
BROW ANACO BROW ANACO PRT; 212 AA.
ID BROW ANACO STANDARD; PRT; 212 AA.
AC P14518;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Bromelain, stem (EC 3.4.22.32).
OS Ananas comosus (Pineapple).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae; Ananas.
OC Bromeliaceae; Ananas.
OX NCBI_TaxID=4615;
RN [1]
RP MEDLINE=89232167; PubMed=2714443;
RX Ritonja A., Rowan A.D., Buttle D.J., Rawlings N.D., Turk V.,
Barrett A.J.;
RT "Stem bromelain: amino acid sequence and implications for weak
binding of cystatin.";
PESS Lett. 247:419-424 (1989).
RN [2]
RP STRUCTURE OF CARBOHYDRATE.
van Kuik J.A., Hoffmann R.A., Mutsaers J.H.G.M., van Halbeek H.,
Kamerling J.P., Vlieghehart J.F.G.;
RT "A 500-MHz 1H-NMR study on the N-linked carbohydrate chain of
bromelain. 1H-NMR structural-reporter-groups of fucose alpha(1-3)-
linked to asparagine-bound N-acetylglucosamine";
Glycoconj. J. 3:27-34 (1986).
RN [3]
CC -!- CATALYTIC ACTIVITY: Broad specificity for cleavage of proteins,
but strong preference for 2-Arg-Arg-|-NHMe amongst small molecule
substrates.
CC -!- MISCELLANEOUS: THE GEOMETRY & THE REACTIVITY OF THE CATALYTIC SITE
ARE DIFFERENT FROM THOSE OF OTHER CYSTEINE PROTEINASES.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
PIR; S03964; S03964.
HSSP; P14080; 1VAL.
MEROPS; C01.005; -.
DR GlycoSuiteDB; P14518; -.
DR InterPro; IPR000668; Peptidase C1.
DR InterPro; IPR000169; SHprot acsite.
DR Pfam; PF00112; Peptidase_C1_1.
DR ProDom; PD000158; Peptidase_C1_1.
DR PROSITE; PS00133; THIOI_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.
DR PROSITE; PS00640; THIOI_PROTEASE_ASN; FALSE_NEG.
DR Hydrolase; Thiol protease; Glycoprotein.
ACT SITE 26 26
ACT SITE 158 158
ACT SITE 23 63
FT DISULFID 23 63

FT DISULFID 57 96
FT DISULFID 152 199
FT CARBOHYD 117 117
FT SEQUENCE 212 AA; 22831 MW; 94EBADB72AAFA556 CRC64;
Query Match 88.6%; Score 93; DB 1; Length 212;
Best Local Similarity 94.7%; Pred. No. 2.7e-08;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSCGAVTSVKNQ 19
DB 2 VPOSIDWRDSCGAVTSVKNQ 20

RESULT 3
PAP2 CARPA
ID _PAP2 CARPA STANDARD; PRT; 352 AA.
AC P14080;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chymopapain precursor (EC 3.4.22.6) (Papaya proteinase II) (PPII).
OS Carica papaya (Papaya).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Caricaceae; Carica.
OX NCBI_TaxID=3649;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RA Connerton I.F.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 135-352.
RX MEDLINE=90179730; PubMed=2106878;
RA Watson D.C., Yaguchi M., Lynn K.R.;
RT "The amino acid sequence of chymopapain from Carica papaya.";
RL Biochem. J. 266:75-81 (1990).
RN [3]
RP SEQUENCE OF 135-352.
RX MEDLINE=89302685; PubMed=2500950;
RA Jacquet A., Kleinschmidt T., Schnek A.G., Looze Y., Braunitzer G.;
RT "The thiol proteinases from the latex of Carica papaya L. III. The
primary structure of chymopapain.";
RL Biol. Chem. Hoppe-Seyler 370:425-434 (1989).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
MEDLINE=97128645; PubMed=8973203;
RA Maes D., Bouckaert J., Poortmans F., Wyns L., Looze Y.;
RT "Structure of chymopapain at 1.7-A resolution.";
RL Biochemistry 35:16292-16298 (1996).
CC -!- CATALYTIC ACTIVITY: Specificity similar to that of papain.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; X97789; CAA66378.1; -.
DR PIR; S04222; S04222.
DR PIR; S08285; S08285.
DR PDB; 1YAL; 23-DEC-96.
DR MEROPS; C01.002; -.
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR000169; SHprot acsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPAIN.
DR ProDom; PD000158; Peptidase_C1; 1.

DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
 KW Hydroxylase; Thiol protease; Zymogen; Signal; 3D-structure.
 FT SIGNAL 1 18
 FT PROPEP 19 134
 FT CHAIN 135 352
 FT ACT_SITE 159 159
 FT ACT_SITE 293 293
 FT ACT_SITE 313 313
 FT DISULFID 156 197
 FT DISULFID 190 229
 FT DISULFID 287 338
 SQ SEQUENCE 352 AA; 39414 MW; 50BA31EBFCFOAF9P CRC64;

Query Match 80.0%; Score 84; DB 1; Length 352;
 Best Local Similarity 84.2%; Pred. No. 1.5e-06;
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 POSIDWRDSCAVTSYKNOG 20
 DB 136 POSIDWRKAGAVTPVKNKG 154

RESULT 4
 ID CATL_PIG STANDARD; PRT; 334 AA.
 AC 028944;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cathepsin L precursor (EC 3.4.22.15).
 GN CTSL.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 OC NCBI_TaxID=9823;
 RX MEDLINE=96069841; PubMed=7492581;
 RA Okamura N., Tamba M., Uchiyama Y., Sugita Y., Dacheux F.,
 RA Syntin P., Dacheux J. L.;
 RT "Direct evidence for the elevated synthesis and secretion of
 RT procathepsin L in the distal caput epididymis of boar.";
 RL Biochim. Biophys. Acta 1245:221-226(1995).
 CC -1- FUNCTION: IMPORTANT FOR THE OVERALL DEGRADATION OF PROTEINS IN
 CC LYSOSOMES.
 CC -1- CATALYTIC ACTIVITY: Specificity close to that of papain. As
 CC compared to cathepsin B, cathepsin L exhibits higher activity
 CC towards protein substrates, but has little activity on Z-Arg-Arg-
 CC NHMeC, and no peptidyl-di-peptidase activity.
 CC -1- SUBUNIT: DIMER OF A HEAVY AND A LIGHT CHAIN LINKED BY DISULFIDE
 CC BONDS.
 CC -1- SUBCELLULAR LOCATION: Lysosomal.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D37917; BA07140.1; -;
 DR HSSP; O60911; IFH0.
 DR MEROPS; C01.032; -;
 DR InterPro; IPR000668; Peptidase_C1.
 DR InterPro; IPR000169; SHprot_acsite.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPAIN.
 DR ProDom; PD000158; Peptidase_C1; 1.

DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
 KW Hydroxylase; Thiol protease; Glycoprotein; Lysosome; Zymogen; Signal.
 FT SIGNAL 1 17
 FT PROPEP 18 117
 FT CHAIN 118 289
 FT PROPEP 290 291
 FT CHAIN 292 334
 FT ACT_SITE 138 138
 FT ACT_SITE 277 277
 FT ACT_SITE 301 301
 FT DISULFID 135 178
 FT DISULFID 169 212
 FT DISULFID 270 323
 FT CARBOHYD 222 222
 FT CARBOHYD 292 292
 SQ SEQUENCE 334 AA; 37178 MW; 51DBA79ACCF2CE53 CRC64;

Query Match 79.0%; Score 83; DB 1; Length 334;
 Best Local Similarity 70.0%; Pred. No. 2.1e-06;
 Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 VPOSIDWRDSCAVTSYKNOG 20
 DB 114 VPKSVDRKREKGYTAVKNOG 133

RESULT 5
 ID CC3_CARCN STANDARD; PRT; 43 AA.
 AC P32956;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cysteine proteinase III (EC 3.4.22.-) (CC-III) (Fragment).
 OS Carica candamarcensis.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;
 OC eurosids II; Brassicales; Caricaceae; Carica.
 OC NCBI_TaxID=29731;
 RX MEDLINE=94030669; PubMed=8216902;
 RA Wairevens V., Jaziri M., van Beeumen J., Schneek A.G.,
 RA Kinschmidt T., Looze Y.;
 RT "Isolation and preliminary characterization of the cysteine-
 RT proteinases from the latex of Carica candamarcensis Hook.";
 RL Biol. Chem. Hoppe-Seyler 374:501-506(1993).
 CC -1- PFM: GLYCOSYLATED.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
 DR HSSP; P14080; IVAL.
 DR MEROPS; C01.020; -;
 DR InterPro; IPR000668; Peptidase_C1.
 DR InterPro; IPR000169; SHprot_acsite.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR ProDom; PD000158; Peptidase_C1; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; PARTIAL.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASN; PARTIAL.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 KW Hydroxylase; Thiol protease; Glycoprotein.
 FT ACT_SITE 25 25
 FT ACT_SITE 301 301
 FT NON_TER 43 43
 SQ SEQUENCE 43 AA; 4636 MW; F4C5D2881886E291 CRC64;

Query Match 78.1%; Score 82; DB 1; Length 43;
 Best Local Similarity 78.9%; Pred. No. 3.2e-07;
 Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 POSIDWRDSCAVTSYKNOG 20
 DB 2 POSIDWRKAGAVTPVKNKG 20

RESULT 6

CC4_CARCIN
ID CC4_CARCIN STANDARD; PRT; 43 AA.
AC P22957;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cysteine proteinase IV (EC 3.4.22.-) (CC-IV) (Fragment).
OS Carica candamarcensis.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eucotyledons II; Brassicales; Caricaceae; Carica.
OX NCBI_TaxID=29731;
RN [1]
RP SEQUENCE.
RC TISSUE=LateX;
RX MEDLINE=94030669; PubMed=8216902;
RA Walravens V., Jaziri M., van Beeumen J., Schneek A.G.,
RA Kleinschmidt T., Looze Y.,
RT "Isolation and preliminary characterization of the cysteine-
RT proteinases from the latex of Carica candamarcensis Hook."
RL Biol. Chem. Hoppe-Seyler 374:501-506 (1993).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
DR HSP; P14080; IYAL.
DR MEROPS; C01.01A; -.
DR InterPro; IPR000668; Peptidase C1.
DR InterPro; IPR000169; SHprot acsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR ProDom; PD000158; Peptidase_C1; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; PARTIAL.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; PARTIAL.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
KW Hydrolase; Thiol protease.
FT ACT_SITE 25 25 BY SIMILARITY.
FT NON_TER 43 43
SQ SEQUENCE 43 AA; 4685 MW; F4D5945A9386E291 CRC64;

Query Match 78.1%; Score 82; DB 1; Length 43;
Best Local Similarity 78.9%; Pred. No. 3.2e-07;

Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 POSIDWRDGSVAVTSVKNQ 20

Db 2 PESIDWRKKGAVTPVKNQ 20

RESULT 7

CATL_MOUSE
ID CATL_MOUSE STANDARD; PRT; 334 AA.
AC P06737;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cathepsin L precursor (EC 3.4.22.15) (Major excreted protein) (MEP).
GN CTSL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=88076849; PubMed=3689328;
RA Troen B.R., Gal S., Gottesman M.M.;
RT "Sequence and expression of the cDNA for MEP (major excreted
RT protein), a transcription-regulated secreted cathepsin.";
RL Biochem. J. 246:731-735 (1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88213715; PubMed=2835398;
RA Joseph L.J., Chang L.C., Stamenkovich D., Sukhatme V.P.;
RT "Complete nucleotide and deduced amino acid sequences of human and

RT murine preprocathepsin L. An abundant transcript induced by
RT transformation of fibroblasts."
RL J. Clin. Invest. 81:1621-1629 (1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=87033683; PubMed=3533924;
RA Portnoy D.A., Erickson A.H., Kochan J., Ravetch J.V., Unkeless J.C.;
RT "Cloning and characterization of a mouse cysteine proteinase.";
RL J. Biol. Chem. 261:14697-14703 (1986).
RN [4]
RP SEQUENCE FROM N.A., AND CARBOHYDRATE-LINKAGE SITE.
RC TISSUE=Liver;
RX MEDLINE=91112761; PubMed=2275556;
RA Stearns N.A., Dong J., Pan J.X., Brenner D.A., Sahagian G.G.;
RT "Comparison of cathepsin L synthesized by normal and transformed
RT cells at the gene, message, protein, and oligosaccharide levels.";
RL Arch. Biochem. Biophys. 283:447-457 (1990).
RN [5]
RP SEQUENCE OF 89-300 FROM N.A.
RC STRAIN=BNL;
RX MEDLINE=86271744; PubMed=3755373;
RA Denhardt D.T., Hamilton R.T., Parfett C.L.J., Edwards D.R.,
RA Pierre R.S., Waterhouse P., Nilson-Hamilton M.;
RT "Close relationship of the major excreted protein of transformed
RT murine fibroblasts to thiol-dependent cathepsins.";
RL Cancer Res. 46:4590-4593 (1986).
CC -1- FUNCTION: IMPORTANT FOR THE OVERALL DEGRADATION OF PROTEINS IN
CC LYSOSOMES.
CC -1- CATALYTIC ACTIVITY. Specificity close to that of papain. As
CC compared to cathepsin B, cathepsin L exhibits higher activity
CC towards protein substrates, but has little activity on Z-Arg-Arg-
CC NHMeC, and no peptidyl-dipeptidase activity.
CC -1- SUBUNIT: DIMER OF A HEAVY AND A LIGHT CHAIN LINKED BY DISULFIDE
CC BONDS.
CC -1- SUBCELLULAR LOCATION: Lysosomal.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X06086; CAA29470.1; -;
CC EMBL; J02583; AAA37445.1; -;
CC EMBL; M20495; AAA39984.1; -;
CC EMBL; X04392; CAA27980.1; -;
CC PIR; S01177; KMSL.
CC PIR; S13890; S13890.
CC HSP; P07711; 1GJL.
CC MEROPS; C01.032; -.
CC MGD; MGI:88564; Ctsl.
CC InterPro; IPR000668; Peptidase_C1.
CC InterPro; IPR000169; SHprot acsite.
CC Pfam; PF00112; Peptidase_C1; 1.
CC PRINTS; PR00705; PAPAIN.
CC ProDom; PD000158; Peptidase_C1; 1.
CC PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
CC PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
CC PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
KW Hydrolase; Thiol protease; Glycoprotein; Lysosome; Zymogen; Signal.
FT SIGNAL 1 17 POTENTIAL;
FT PROPEP 18 113 ACTIVATION PEPTIDE.
FT CHAIN 114 288 CATHEPSIN L, HEAVY CHAIN.
FT CHAIN 291 334 CATHEPSIN L, LIGHT CHAIN.
FT ACT_SITE 138 138 BY SIMILARITY.
FT ACT_SITE 276 276 BY SIMILARITY.
FT ACT_SITE 300 300 BY SIMILARITY.
FT DISULFD 135 178 BY SIMILARITY.
FT DISULFD 169 211 BY SIMILARITY.
FT DISULFD 269 322 INTERCHAIN (BY SIMILARITY).

FT CARBOHYD 221 221 N-LINKED (GLCNAC...)
 FT CONFLICT 58 58 M -> I (IN REF. 2).
 FT CONFLICT 177 177 G -> R (IN REF. 3).
 SO SEQUENCE 334 AA; 37547 MW; FE674704307AD98 CRC64;

Query Match 78.1%; Score 82; DB 1; Length 334;
 Best Local Similarity 65.0%; Pred. No. 3e-06;
 Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPOSIDWRDGAATSVKNKG 20
 Db 114 IPKSVPMREKGCCTPVKNKG 133

RESULT 8
 ID CATL_SCHEMA STANDARD; PRT; 319 AA.
 AC Q26534;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUN-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cathepsin L precursor (EC 3.4.22.15) (SMC11).
 GN C11.
 OS Schistosoma mansoni (Blood fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
 OC Schistosomatidae; Schistosomatidae; Schistosoma.
 NC NCB1_TaxID=6183;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Puerto Rican;
 RX MEDLINE=95140043; PubMed=7838171;
 RA Smith A.M., Dalton J.P., Clough K.A., Kilbane C.L., Harrop S.A.,
 RA Hole N., Brindley P.J.;
 RT "Adult Schistosoma mansoni express cathepsin L proteinase activity."
 RL Mol. Biochem. Parasitol. 67:11-19(1994).
 CC -1- FUNCTION: MAY BE CRUCIAL FOR METABOLISM OF HOST HEMOGLOBIN.
 CC -1- CATALYTIC ACTIVITY: Specificity close to that of papain. As
 CC compared to cathepsin B, cathepsin L exhibits higher activity
 CC towards protein substrates, but has little activity on Z-Arg-Arg-
 CC NMeC, and no peptidyl-dipeptidase activity.
 CC -1- SUBCELLULAR LOCATION: Lysosomal.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U07345; AAC6485.1; -
 DR HSSP; O60911; IFHO.
 DR MEROPS; C01.018; -
 DR InterPro; IPR000668; Peptidase_C1.
 DR InterPro; IPR000169; SHprot_acsite.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPA1N.
 DR PRODOM; PD000158; Peptidase_C1; 1.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
 KM Hydroxylase; Thiol protease; Glycoprotein; Lysosome; Zymogen; Signal.
 FT SIGNAL 1
 FT PROPEP 1
 FT CHAIN 105 104
 FT ACT_SITE 129 129 CATHESPIN L.
 FT ACT_SITE 265 265 BY SIMILARITY.
 FT ACT_SITE 286 286 BY SIMILARITY.
 FT DISULFID 126 167 BY SIMILARITY.
 FT DISULFID 160 200 BY SIMILARITY.
 FT DISULFID 258 307 BY SIMILARITY.
 SO SEQUENCE 319 AA; 36136 MW; 9CFF6A3F1193479 CRC64;

Query Match 77.1%; Score 81; DB 1; Length 319;
 Best Local Similarity 65.0%; Pred. No. 4.2e-06;
 Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPOSIDWRDGAATSVKNKG 20
 Db 105 IPKSVPMREKGCCTPVKNKG 124

RESULT 9
 ID A494_ARATH STANDARD; PRT; 361 AA.
 AC P43255; O9SUT5;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable cysteine proteinase A494 precursor (EC 3.4.22.-).
 GN AT2G21430 OR P3K23.19.
 GN Arabidopsis thaliana (Mouse-ear cress).
 OS Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucoside II; Brassicales; Brassicaceae; Arabidopsids.
 NC NCB1_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
 RA Beffell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
 RA Boffett C.S., Cronin L.A., Shen M., Pal G., Van Aken S., Umayam L.,
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
 RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
 RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
 RA Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 RT thaliana." Nature 402:761-768(1999).
 RL Nature 402:761-768(1999).
 RN [2]
 RP SEQUENCE OF 49-361 FROM N.A.
 RC STRAIN=cv. Landsberg erecta;
 RX MEDLINE=94289649; PubMed=8018874;
 RA Williams J., Bulman M., Huttly A.K., Phillips A., Neill S.;
 RT "Characterization of a cDNA from Arabidopsis thaliana encoding a
 RT potential thiol protease whose expression is induced independently by
 RT wounding and abscisic acid." Plant Mol. Biol. 25:259-270(1994).
 CC -1- INDUCTION: BY WOUNDING AND ABSICISIC ACID (ABA).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AC006841; AAD23687.1; -
 DR EMBL; X74359; CA52403.1; -
 DR HSSP; P07711; 1CUL.
 DR MEROPS; C01.022; -
 DR InterPro; IPR000668; Peptidase_C1.
 DR InterPro; IPR000169; SHprot_acsite.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPA1N.
 DR PRODOM; PD000158; Peptidase_C1; 1.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
 KM Hydroxylase; Thiol protease; Zymogen; Signal.
 FT SIGNAL 1
 FT PROPEP 23
 FT CHAIN 23
 FT ACT_SITE 129 129 CATHESPIN L.
 FT ACT_SITE 265 265 BY SIMILARITY.
 FT ACT_SITE 286 286 BY SIMILARITY.
 FT DISULFID 126 167 BY SIMILARITY.
 FT DISULFID 160 200 BY SIMILARITY.
 FT DISULFID 258 307 BY SIMILARITY.
 SO SEQUENCE 319 AA; 36136 MW; 9CFF6A3F1193479 CRC64;

FT PROPEP 24 131 ACTIVATION PEPTIDE (POTENTIAL)
 FT CHAIN 132 361 PROBABLE CYSTEINE PROTEINASE A494.
 FT ACT_SITE 136 156 BY SIMILARITY.
 FT ACT_SITE 299 299 BY SIMILARITY.
 FT ACT_SITE 326 326 BY SIMILARITY.
 FT DISULFID 153 203 BY SIMILARITY.
 FT DISULFID 187 237 BY SIMILARITY.
 FT DISULFID 293 347 BY SIMILARITY.
 FT CONFLICT 49 49 T -> A (IN REF. 2).
 SQ SEQUENCE 361 AA; 39819 MW; D09A2CE564642DD CRC64;
 Query Match 77.1%; Score 81; DB 1; Length 361;
 Best Local Similarity 70.0%; Pred. No. 4.9e-06;
 Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 VPOSIDWRDGSVATSVKNOG 20
 Db 132 LPDEFDRDGRGAVTPVKNOG 151
 RESULT 10
 RD19 ARATH STANDARD; PRT; 368 AA.
 AC P43296;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cysteine proteinase RD19A precursor (EC 3.4.22.-).
 GN RD19A OR A14G39090 OR F19H22.190.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=93314960; PubMed=8325504;
 RA Kohzumi M., Yamaguchi-Shinozaki K., Tsuji H., Shinozaki K.;
 FT "Structure and expression of two genes that encode distinct drought-
 RT inducible cysteine proteinases in Arabidopsis thaliana.";
 RL Gene 129:175-182 (1993).
 SQ SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,
 RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
 RA Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidtheini T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos F., Hohnheisel J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
 RA Vander Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
 RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Weitzenecker T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
 RA Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
 RA Bernier S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
 RA De Keyser A., Buysshaert C., Gielens J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
 RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rehnann S.,
 RA Borkova D., Bloeker H., Scharf M., Grimm M., Loehert T.-H.,
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Farmann B., Granderath K., Dauner D., Herzl A.,
 RA Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
 RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Chedor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,

RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Biele C.,
 RA Frisman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dethia N., Ghoj L., Schutz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Sheet P., Cordes K., Abu-Threideh J.,
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton L., Miller N., Greco T., Kemp K.,
 RA Kramer J., Spith J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Anonou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shan R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
 RA Chen E., Marra M., Martienssen R., McCombie W.R.;
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 thaliana.";
 RL Nature 402:769-777 (1999).
 CC -!- INDUCTION: BY HIGH SALT CONDITIONS.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL curators.
 CC The European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC EMBL; D13042; BAA02373.1; -;
 DR EMBL; AL035679; CAB38829.1; -;
 DR EMBL; AL161594; CAB80572.1; -;
 DR KSSP; P07711; 1CJL.
 DR MEROPS; C01.022; -;
 DR InterPro; IPR000668; Peptidase C1.
 DR InterPro; IPR000169; SHprot acSite.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPAIN_C1; 1.
 DR ProDom; PD000158; Peptidase C1; 1.
 DR ProSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 DR ProSITE; PS00639; THIOL_PROTEASE_HIS; 1.
 DR ProSITE; PS00640; THIOL_PROTEASE_ASN; 1.
 KW Hydrolase; Thiol protease; Zymogen; Glycoprotein; Signal.
 FT SIGNAL 1 23 POTENTIAL.
 FT PROPEP 24 134 ACTIVATION PEPTIDE (POTENTIAL).
 FT CHAIN 135 368 CYSTEINE PROTEINASE RD19A.
 FT ACT_SITE 159 159 BY SIMILARITY.
 FT ACT_SITE 302 302 BY SIMILARITY.
 FT ACT_SITE 329 329 BY SIMILARITY.
 FT DISULFID 156 206 BY SIMILARITY.
 FT DISULFID 194 240 BY SIMILARITY.
 FT DISULFID 296 350 BY SIMILARITY.
 FT CARBOHYD 253 253 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 368 AA; 40418 MW; 9D64CF35E07F519D CRC64;
 Query Match 77.1%; Score 81; DB 1; Length 368;
 Best Local Similarity 70.0%; Pred. No. 4.9e-06;
 Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 VPOSIDWRDGSVATSVKNOG 20
 Db 135 LPDEFDRDGRGAVTPVKNOG 154
 RESULT 11
 GP11_ZINOF STANDARD; PRT; 221 AA.
 ID GP11_ZINOF
 AC P82474;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cysteine proteinase GP-II (EC 3.4.22.-).

OS Zingiber officinale (Ginger).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Zingiberales; Zingiberaceae;
OC Zingiber.
OX NCBI_TaxID=94328;
RN [1]
RP SEQUENCE.
RC TISSUE=Root;
RX MEDLINE=20156257; PubMed=10691991;
RA Choi K.H., Laursen R.A.;
RT "Amino-acid sequence and glycan structures of cysteine proteases with
RT proline specificity from ginger rhizome Zingiber officinale.",
RL Eur. J. Biochem. 267:1516-1526(2000).
CC -1- CATALYTIC ACTIVITY: Preferential cleavage with a proline residue
CC at P2.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
DR HSSP; P00785; 2ACT.
DR GLYCOSITE; P82474; -.
DR InterPro: IPR000668; SHPoc_acsite.
DR Pfam: PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPA1N.
DR ProDom; PD000158; Peptidase_C1; 1.
DR PROSITE; PS00139; THIOL_PROTEASE_Cys; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_His; 1.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
KM Hydrolyase; Thiol protease; Glycoprotein.
FT ACT_SITE 27 BY SIMILARITY.
FT ACT_SITE 161 BY SIMILARITY.
FT DISULFID 24 BY SIMILARITY.
FT DISULFID 58 BY SIMILARITY.
FT DISULFID 155 BY SIMILARITY.
FT CARBOHYD 99 N-LINKED (GLCNAC. . .).
FT CARBOHYD 156 N-LINKED (GLCNAC. . .).
FT SEQUENCE 221 AA; 23922 MW; 909A312BD8632D42 CRC64;
Query Match 76.2%; Score 80; DB 1; Length 221;
Best Local Similarity 70.0%; Pred. No. 4.2e-06;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
Oy 1 VPOSIDRWDSGAVTSYKNOG 20
Db 3 LPDSIDRWENGAVPYKNOG 22
RESULT 12
ID ORYA_ORYSA STANDARD; PRT; 471 AA.
AC P25777;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Oryza beta chain precursor (EC 3.4.22.-).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eriaristaceae; Oryzae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare; TISSUE=Seed;
RX MEDLINE=91358494; PubMed=1885617;
RA Watanabe H., Abe K., Emori Y., Hosoyama H., Arai S.;
RT "Molecular cloning and gibberellin-induced expression of multiple
RT cysteine proteinases of rice seeds (oryza sativa).",
RL J. Biol. Chem. 266:16897-16902(1991).
CC -1- TISSUE SPECIFICITY: EXPRESSED ONLY IN SEEDS.
CC -1- INDUCTION: BY GIBBERELLIC ACID (GA).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90407; BAA14403.1; -.
DR PIR; J00389; KHRZOB.
DR HSSP; P00785; 2ACT.
DR MEROPS; C01.029; -.
DR InterPro: IPR000118; Granulin.
DR InterPro: IPR000668; Peptidase_C1.
DR InterPro: IPR00169; SHPoc_acsite.
DR Pfam; PF00396; granulin; 1.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPA1N.
DR ProDom; PD000158; Peptidase_C1; 1.
DR SMART; SM00277; GRAN.1.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR PROSITE; PS00139; THIOL_PROTEASE_Cys; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_His; 1.
KM Hydrolyase; Thiol protease; Zymogen; Glycoprotein; Signal.
FT SIGNAL 1 21
FT PROPEP 22 139 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 140 471 ORYZAIN BETA CHAIN.
FT ACT_SITE 164 164 BY SIMILARITY.
FT ACT_SITE 301 301 BY SIMILARITY.
FT ACT_SITE 321 321 BY SIMILARITY.
FT DISULFID 161 204 BY SIMILARITY.
FT DISULFID 195 237 BY SIMILARITY.
FT DISULFID 295 346 BY SIMILARITY.
FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 471 AA; 50505 MW; 08B39D3395CEB1D CRC64;
Query Match 76.2%; Score 80; DB 1; Length 471;
Best Local Similarity 65.0%; Pred. No. 9.5e-06;
Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
Oy 1 VPOSIDRWDSGAVTSYKNOG 20
Db 140 LPESVDWRKGVAPYKNOG 159
RESULT 13
ID GPI_ZINOF STANDARD; PRT; 221 AA.
AC P82473;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cysteine proteinase GP-I (EC 3.4.22.-).
OS Zingiber officinale (Ginger).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Zingiberales; Zingiberaceae;
OC Zingiber.
OX NCBI_TaxID=94328;
RN [1]
RP SEQUENCE.
RC TISSUE=Root;
RX MEDLINE=20156257; PubMed=10691991;
RA Choi K.H., Laursen R.A.;
RT "Amino-acid sequence and glycan structures of cysteine proteases with
RT proline specificity from ginger rhizome Zingiber officinale.",
RL Eur. J. Biochem. 267:1516-1526(2000).
CC -1- CATALYTIC ACTIVITY: Preferential cleavage with a proline residue
CC at P2.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
CC -1- CAUTION: THE AUTHORS REGARD THE SEQUENCE AS TENTATIVE, AS THEY
CC BELIEVE THAT IT MAY HAVE BEEN CONTAMINATED BY A HOMOLOGOUS
CC PROTEIN.
DR HSSP; P00785; 2ACT.

DR InterPro; IPR000668; Peptidase Cl.
 DR InterPro; IPR000169; SHprot_acsite.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPAIN.
 DR ProDom; PD000158; Peptidase_C1; 1.
 DR PROSITE; PS00139; THIOI_PROTEASE_CYS; FALSE_NEG.
 DR PROSITE; PS00639; THIOI_PROTEASE_HIS; FALSE_NEG.
 DR PROSITE; PS00640; THIOI_PROTEASE_ASN; 1.
 KW Hydrolase; Thiol protease; Glycoprotein.
 FT ACT SITE 27 27 BY SIMILARITY.
 FT ACT SITE 161 161 BY SIMILARITY.
 FT DISULFID 24 65 BY SIMILARITY.
 FT DISULFID 58 98 BY SIMILARITY.
 FT DISULFID 155 206 BY SIMILARITY.
 FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 156 156 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT UNSURE 2 2 OR D.
 SQ SEQUENCE 221 AA; 24241 MW; 3035D7870EA743DB CRC64;
 Query Match 75.2%; Score 79; DB 1; Length 221;
 Best Local Similarity 70.0%; Pred. No. 6.1e-06;
 Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 VPQSIDWRDSCAVTSVKNQ 20
 : |||||: |||||
 Db 3 LPDSIDMRKGVVVPKNQ 22
 RESULT 14
 CATL RAT STANDARD; PRT; 334 AA.
 AC F07154; QSQV07;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cathepsin L precursor (EC 3.4.22.15) (Major excreted protein) (MEP)
 DE (Cyclic protein-2) (CP-2).
 GN CTSL.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wistar; TISSUE=Kidney;
 RX MEDLINE=88030047; PubMed=3666143;
 RA Ishido K., Towatari T., Imajoh S., Kawasaki H., Kominami E.,
 RA Katunuma N., Suzuki K.;
 RT "Molecular cloning and sequencing of cDNA for rat cathepsin L.";
 RL FEBS Lett. 223:69-73(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90092543; PubMed=2599113;
 RA Ishido K., Kominami E., Suzuki K., Katunuma N.;
 RT "Gene structure and 5'-upstream sequence of rat cathepsin L.";
 RL FEBS Lett. 259:71-74(1989).
 RN [3]
 RP SEQUENCE OF 1-42 FROM N.A., TISSUE SPECIFICITY, AND INDUCTION.
 RC STRAIN=Sprague-Dawley;
 RX MEDLINE=21255611; PubMed=11356678;
 RA Zabludoff S.D., Charron M., DeCarbo J.N., Simukova N., Wright W.W.;
 RT "Male germ cells regulate transcription of the cathepsin L gene by rat Sertoli cells.";
 RL Endocrinology 142:2318-2327(2001).
 RN [4]
 RP SEQUENCE OF 88-334 FROM N.A.
 RC TISSUE=Sertoli cells;
 RX MEDLINE=92168015; PubMed=1791830;
 RA Erickson-Lawrence M., Zabludoff S.D., Wright W.W.;
 RT "Cyclic protein-2, a secretory product of rat Sertoli cells, is the proenzyme form of cathepsin L.";
 RL Mol. Endocrinol. 5:1789-1798(1991).
 RN [5]

RP SEQUENCE OF 18-37, FUNCTION, AND SUBCELLULAR LOCATION.
 RC STRAIN=Sprague-Dawley; TISSUE=Sertoli cells;
 RX MEDLINE=95296691; PubMed=7777858;
 RA Boujrad N., Oguegbu S.O., Garnier M., Lee C.-H., Martin B.M.,
 RA Papadopoulos V.;
 RT "Identification of a stimulator of steroid hormone synthesis isolated from testis.";
 RL Science 268:1609-1612(1995).
 RN [6]
 RP SEQUENCE OF 18-28, AND TISSUE SPECIFICITY.
 RC STRAIN=Sprague-Dawley; TISSUE=Epidermis;
 RX MEDLINE=20164186; PubMed=10699763;
 RA Kawada A., Hara K., Kominami E., Tezuka T., Takahashi M., Takahara H.;
 RT "Precursor of rat epidermal cathepsin L: purification and immunohistochemical localization.";
 RL J. Dermatol. Sci. 23:36-45(2000).
 RN [7]
 RP SEQUENCE OF 114-288 AND 291-334.
 RC TISSUE=Liver;
 RX MEDLINE=88296890; PubMed=3402618;
 RA Towatari T., Katunuma N.;
 RT "Amino acid sequence of rat liver cathepsin L.";
 RL FEBS Lett. 236:57-61(1988).
 CC -!- FUNCTION: Important for the overall degradation of proteins in lysosomes. Procathepsin L is required for maximal stimulation of steroidogenesis by TIMP1.
 CC -!- CATALYTIC ACTIVITY: Specificity close to that of papain. As compared to cathepsin B, cathepsin L exhibits higher activity towards protein substrates but has little activity on Z-Arg-Arg-NHec, and no peptidyl-dipeptidase activity.
 CC -!- SUBUNIT: DIMER OF A HEAVY AND A LIGHT CHAIN LINKED BY DISULFIDE BONDS.
 CC -!- SUBCELLULAR LOCATION: Procathepsin L is secreted. Mature cathepsin L is lysosomal.
 CC -!- TISSUE SPECIFICITY: Both mature cathepsin L and procathepsin L are found in the upper epidermis. The lower epidermis predominantly contains procathepsin L. In seminiferous tubules expression is greater at stages VI-VII than at stages IX-XII.
 CC -!- INDUCTION: Expression in Sertoli cells is repressed by germ cells.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 DR EMBL; Y00697; CAA68691.1; -;
 DR EMBL; AF025476; AAB81616.1; -;
 DR EMBL; S85184; AAB21516.1; -;
 DR PIR; S07098; KHRTL.
 DR HSSP; O60911; IFHO.
 DR MEROPS; C01.032; -;
 DR InterPro; IPR000668; Peptidase_C1.
 DR InterPro; IPR000169; SHprot_acsite.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPAIN.
 DR ProDom; PD000158; Peptidase_C1; 1.
 DR PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.
 DR PROSITE; PS00640; THIOI_PROTEASE_ASN; 1.
 KW Hydrolase; Thiol protease; Glycoprotein; Lysosome; Zymogen; Signal.
 FT SIGNAL 1 17
 FT PROPEP 18 113 ACTIVATION PEPTIDE.
 FT CHAIN 114 288 CATHEPSIN L, HEAVY CHAIN.
 FT CHAIN 291 334 CATHEPSIN L, LIGHT CHAIN.
 FT ACT SITE 138 138 BY SIMILARITY.
 FT ACT_SITE 276 276 BY SIMILARITY.
 FT ACT_SITE 300 300 BY SIMILARITY.
 FT DISULFID 135 178 BY SIMILARITY.
 FT DISULFID 169 211 BY SIMILARITY.

FT DISULFID 269 322 INTERCHAIN (BY SIMILARITY).
 FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .).
 FT CONFLICT 238 238 A -> P (IN REF. 1).
 SQ SEQUENCE 334 AA; 37660 MM; AFA997582E34AF6 CRC64;
 Query Match 75.2%; Score 79; DB 1; Length 334;
 Best Local Similarity 60.0%; Pred. NO. 9.6e-06;
 Matches 12; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
 QY 1 VPOSIDMRDSGAVTSVXNOG 20
 Db 114 IPRTVDMREKGCVTPEKNOG 133
 RESULT 15
 ID PAPA_CARPA STANDARD; PRT; 345 AA.
 AC P00784;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Papain precursor (EC 3.4.22.2) (papaya proteinase I) (PPI).
 OS Carica papaya (Papaya).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Caricaceae; Carica.
 OC NCBI_TaxID=3649;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87163523; PubMed=2881845;
 RA Cohen L.W., Coghlan V.M., Dibel L.C.;
 RT "Cloning and sequencing of papain-encoding cDNA.";
 RL Gene 48:219-227(1986).
 RN [2]
 RP SEQUENCE OF 134-345.
 RX MEDLINE=71007899; PubMed=5470818;
 RA Mitchell R.E.J., Chaiken I.M., Smith E.L.;
 RT "The complete amino acid sequence of papain. Additions and
 RT corrections.";
 RL J. Biol. Chem. 245:3485-3492(1970).
 RN [3]
 RP REVISION TO 197.
 RX MEDLINE=70141125; PubMed=5435495;
 RA Husain S.S., Lowe G.;
 RT "A reinvestigation of residues 64-68 and 175 in papain. Evidence that
 RT residues 64 and 175 are asparagine.";
 RL Biochem. J. 116:689-692(1970).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 RX MEDLINE=6906973; PubMed=5681232;
 RA Drenth J., Jansonsius J.N., Koekoek R., Swen H.M., Wolthers B.G.;
 RT "Structure of papain.";
 RL Nature 218:929-932(1968).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).
 RX MEDLINE=85058190; PubMed=6502713;
 RA Kamphuis I.G., Kalk K.H., Swarte M.B.A., Drenth J.;
 RT "Structure of papain refined at 1.65-A resolution.";
 RL J. Mol. Biol. 179:233-256(1984).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 RX MEDLINE=90269230; PubMed=2347312;
 RA Stubbs M.T., Laber B., Bode W., Huber R., Jerala R., Lenarcic B.,
 RA Turk V.;
 RT "The refined 2.4 A X-ray crystal structure of recombinant human
 RT stefin B in complex with the cysteine proteinase papain: a novel type
 RT of proteinase inhibitor interaction.";
 RL EMBO J. 9:1939-1947(1990).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
 RX MEDLINE=93075728; PubMed=1445868;
 RA Yamamoto K., Tomoo K., Doi M., Onishi H., Inoue M., Ishida T.,
 RA Yamamoto D., Tsuboi S., Okamoto H., Okada Y.;

RT "Crystal structure of
 RT papain-succinyl-Gln-Val-Ala-Ala-p-nitroanilide complex at 1.7-A
 RT resolution: noncovalent binding mode of a common sequence of
 RT endogenous thiol protease inhibitors.";
 RL Biochemistry 31:11305-11309(1992).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
 RA Pickersall R.W., Harris G.W., Garman E.;
 RT "Structure of monoclinic papain at 1.60-A resolution.";
 RL Acta Crystallogr. B 48:59-67(1992).
 CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-, PHE-XAA-
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
 CC -1- DATABASE: NAME=Worthington enzyme manual;
 CC WWW="http://www.worthington-biochem.com/manual/P/PAP.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M15203; AAB02650.1; .
 DR PIR; A26466; PPA.
 DR PDB; 1PAD; 17-FEB-84.
 DR PDB; 2PAD; 17-FEB-84.
 DR PDB; 3PAD; 04-MAR-85.
 DR PDB; 5PAD; 17-FEB-84.
 DR PDB; 6PAD; 17-FEB-84.
 DR PDB; 9PAD; 15-JAN-95.
 DR PDB; 1P6; 15-APR-93.
 DR PDB; 1P1P; 31-OCT-93.
 DR PDB; 1P0P; 31-OCT-93.
 DR PDB; 1PPD; 02-JAN-85.
 DR PDB; 1PPN; 31-JAN-94.
 DR PDB; 1PPP; 31-JAN-94.
 DR PDB; 1STF; 31-JAN-94.
 DR MEROPS; C01.001; .
 DR InterPro; IPR000668; Peptidase C1.
 DR InterPro; IPR000169; SH3prot acsite.
 DR Pfam; PF00112; Peptidase_C1.1.
 DR PRINTS; PRO0705; PAPAIN_C1.1.
 DR ProDom; PD000158; Peptidase_C1.1.
 DR ProSITE; PS00139; THIOL_PROTEASE_CYS.1.
 DR ProSITE; PS00639; THIOL_PROTEASE_HIS.1.
 DR ProSITE; PS00640; THIOL_PROTEASE_ASN.1.
 KW Hydroxylase; Thiol protease; Zymogen; Signal; 3D-structure.
 FT SIGNAL 1
 FT PROPEP 19
 FT CHAIN 134 345
 FT ACT_SITE 158 159
 FT ACT_SITE 292 292
 FT ACT_SITE 308 308
 FT DISULFID 155 196
 FT DISULFID 189 228
 FT DISULFID 286 333
 FT CONFLICT 180 180
 FT CONFLICT 219 180
 FT CONFLICT 251 220
 FT CONFLICT 251 251
 FT CONFLICT 268 268
 FT STRAND 138 139
 FT TURN 140 144
 FT TURN 151 151
 FT TURN 153 154
 FT STRAND 156 156
 FT HELIX 158 175
 FT STRAND 181 181
 FT HELIX 183 189
 FT TURN 191 192
 FT TURN 195 196
 FT STRAND 197 197
 FT STRAND 199 199
 E -> Q (IN REF. 2).
 YP -> PY (IN REF. 2).
 E -> Q (IN REF. 2).
 E -> Q (IN REF. 2).
 E -> Q (IN REF. 2).

FT HELIX 201 210
 FT TURN 211 211
 FT STRAND 213 213
 FT STRAND 215 215
 FT TURN 216 218
 FT HELIX 231 233
 FT STRAND 238 238
 FT STRAND 242 245
 FT HELIX 251 260
 FT STRAND 263 267
 FT HELIX 272 276
 FT STRAND 281 282
 FT STRAND 292 299
 FT STRAND 303 307
 FT STRAND 310 310
 FT TURN 312 313
 FT TURN 315 315
 FT STRAND 316 316
 FT TURN 317 318
 FT STRAND 319 323
 FT HELIX 332 334
 FT TURN 335 336
 FT STRAND 340 343
 SQ SEQUENCE 345 AA; 38922 MW; 82D9FB35EDCA12EF CRC64;

Query Match 75.2%; Score 79; DB 1; Length 345;
 Best Local Similarity 65.0%; Pred. NO. 9.9e-06;
 Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VPQSIDWRDSCGAVTSKNOG 20
 Db 134 IPEYVDMRQKGAVTPKNOG 153

Search completed: November 21, 2002, 16:09:02
 Job time : 8.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2002, 15:59:48 : Search time 7.5 Seconds

(without alignments)
110.604 Million cell updates/sec

Title: US-09-674-738-1

Perfect score: 108

Sequence: 1 VPQSDWRDYGAVTSVKNN 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seque, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	108	100.0	212 1 BROM_ANACO	P14518 ananas como
2	93	86.1	345 1 ANAN_ANACO	P00884 ananas como
3	78	72.2	368 1 RD19_ARATH	P43296 arabidopsis
4	76	70.4	352 1 PAP2_CARPA	P14080 carica papa
5	75	69.4	334 1 CATL_PIG	P28994 sus scrofa
6	74	68.5	43 1 CC3_CARCN	P32956 carica cand
7	74	68.5	43 1 CC4_CARCN	P32957 carica cand
8	74	68.5	334 1 CATL_MOUSE	P06797 mus musculus
9	73	68.5	361 1 A494_ARATH	P43295 arabidopsis
10	73	67.6	319 1 CATL_SCHMA	P25834 schistocoma
11	72	66.7	19 1 CAT3_FASHE	P80533 fasciola he
12	72	66.7	471 1 ORYA_ORYSA	P25777 oryza sativ
13	71	65.7	43 1 CC2_CARCN	P32955 carica cand
14	71	65.7	221 1 GP1_ZINOF	P82474 zingiber of
15	71	65.7	221 1 GP1_ZINOF	P82473 zingiber of
16	71	65.7	334 1 CATL_PAP	P07154 rattus norv
17	71	65.7	345 1 PAPA_CARPA	P00784 carica papa
18	71	65.7	354 1 CYS1_LEIPT	P35591 leishmania
19	71	65.7	354 1 LCPA_LEIME	P25804 leishmania
20	70	64.8	341 1 CATL_DROME	P25802 drosophila
21	70	64.8	362 1 CYP_PHAUV	P25803 phaseolus v
22	70	64.8	362 1 CYP_VIGMU	P12412 vigna mungo
23	70	64.8	443 1 LCPA_LEIME	P26400 leishmania
24	70	64.8	444 1 CYS2_LEIPT	P05099 leishmania
25	69	63.9	333 1 CATL_HUMAN	P07711 homo sapien
26	69	63.9	371 1 CYS1_MAIZE	P10718 zea mays (m
27	68	63.0	348 1 PAPA_CARPA	P00784 carica papa
28	68	63.0	371 1 CYS1_HORVU	P25249 hordeum vul
29	68	63.0	373 1 CYS2_HORVU	P25250 hordeum vul
30	67	62.0	333 1 TEST1_RAT	P15244 rattus norv
31	67	62.0	334 1 CS12_HUMAN	P06091 homo sapien
32	67	62.0	343 1 CYS1_DICDI	P04988 dictyoseli
33	67	62.0	346 1 CATV_GYXN	P09953 xestia c-ni

34	66	61.1	462 1 RD21_ARATH	P43297 arabidopsis
35	65	60.2	330 1 CATS_RAT	P02765 rattus norv
36	65	60.2	333 1 CATV_GVCP	P09166 cydia pomon
37	65	60.2	363 1 CYP_PEA	P25804 pisum sativ
38	65	60.2	395 1 CATL_BRUPA	P17473 brugia paba
39	64	59.3	217 1 CATL_SHEEP	P10991 ovis aries
40	64	59.3	218 1 CATL_CHICK	P09648 gallus gall
41	64	59.3	331 1 CATS_HUMAN	P25774 homo sapien
42	64	59.3	334 1 CATL_BOVIN	P25975 bos taurus
43	64	59.3	458 1 ORYA_ORYSA	P25776 oryza sativ
44	63	58.3	43 1 CCI_CARCN	P32954 carica cand
45	63	58.3	348 1 PAP3_CARPA	P10056 carica papa

ALIGNMENTS

RESULT 1
BROM_ANACO STANDARD; PRT; 212 AA.
ID BROM_ANACO
AC P14518;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Bromelain, stem (EC 3.4.22.32).
OS Ananas comosus (Pineapple).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae; Insectae sedis;
OC Bromeliaceae; Ananas.
OX NCBI_TaxID=4615;
RN [1]
RP SEQUENCE.
RX MEDLINE=89232167; PubMed=2714443;
RA Ralston A., Rowan A.D., Buttle D.J., Rawlings N.D., Turk V.,
RA Barrett A.J.;
RT "Stem bromelain: amino acid sequence and implications for weak
RT binding of cystatin.",
RL FEBS Lett. 247:419-424 (1989).
RN [2]
RP STRUCTURE OF CARBOHYDRATE.
RA van Kuik J.A., Hoffmann R.A., Mutsaers J.H.G.M., van Halbeek H.,
RA Kamerling J.P., Vliegenhart J.F.G.;
RT "A 500-MHz 1H-NMR study on the N-linked carbohydrate chain of
RT bromelain. 1H-NMR structural-reporter-groups of fucose alpha(1-3)-
RT linked to asparagine-bound N-acetylglucosamine.",
RL Glycoconj. J. 3:27-34 (1986).
CC -1- CATALYTIC ACTIVITY: Broad specificity for cleavage of proteins,
CC but strong preference for Z-Arg-Arg-|-NHec amongst small molecule
CC substrates.
CC -1- MISCELLANEOUS: THE GEOMETRY & THE REACTIVITY OF THE CATALYTIC SITE
CC ARE DIFFERENT FROM THOSE OF OTHER CYSTEINE PROTEINASES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
CC PIR: S03964; S03964.
DR HSSP: P14080; 1YAL.
DR MEROPS: C01.005; .
DR GLYCOSULEDB: P14518; .
DR InterPro: IPR000668; Peptidase_C1.
DR InterPro: IPR000169; SHprot_acsite.
DR Pfam: PF00112; Peptidase_C1; 1.
DR ProDom: PD000158; Peptidase_C1; 1.
DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE: PS00639; THIOL_PROTEASE_HIS; 1.
DR PROSITE: PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
DR Hydrobase: Thiol protease; Glycoprotein
KW ACT_SITE; 26
FT ACT_SITE 26
FT ACT_SITE 158
FT DISULFID 23
FT DISULFID 57
FT DISULFID 96
FT DISULFID 152
FT CARBOHYD 117
FT SEQUENCE 212 AA; 22831 MW; 94EBADB72AARV556 CRC64;

```

Query Match          100.0%; Score 108; DB 1; Length 212;
Best Local Similarity 100.0%; Pred. No. 2.5e-10; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0;

QY 1 VPOSIDWRDYGAVTSVKQN 20
    |||||
Db 2 VPOSIDWRDYGAVTSVKQN 21
    |||||

RESULT 2
ANAN ANACO
ID ANAN ANACO STANDARD; PRT; 345 AA.
AC P00884; O22293;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ananain precursor (EC 3.4.22.31).
GN ANI.
OS Ananas comosus (Pineapple).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae incertae sedis;
OC Bromeliaceae; Ananas.
OX NCBI_TaxID=4615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Smooth Cayenne; TISSUE=stem;
RA Robertson C.E., Goodenough P.W.;
RT "Cloning and expression of ananain gene from pineapple.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 123-338.
RC TISSUE=stem;
RX MEDLINE=97454461; PubMed=9355753;
RA Lee K.L., Albee K.L., Bernasconi R.J., Edmunds T.;
RT "Complete amino acid sequence of ananain and a comparison with stem
bromelain and other plant cysteine proteases.";
RL Biochem. J. 327:199-202(1997).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
for peptide bonds. Best reported small molecule substrate Bz-Phe-
Val-Arg-|-NHMe, but broader specificity than fruit bromelain.
CC -!- ENZYME REGULATION: STRONGLY INHIBITED BY CHICKEN EGG-WHITE
CYSTATIN AND THE ACTIVE-SITE-DIRECTED INHIBITOR TRANS-
EPHOXYSUCCINYL-L-LEUCYLAMIDO-(4-GUANIDINO)BUTANE (E-64).
CC -!- TISSUE SPECIFICITY: STEM.
CC -!- MASS SPECTROMETRY: MW=23478; METHOD=Electrospray.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AJ002477; CAA05487.1; -
CC HSPB; P14080; IVAL.
CC MEROPS; C01.026; -.
CC InetPro; IPR000668; Peptidase_C1.
CC InetPro; IPR000169; SHprot_acs1ce.
CC Fram; PF00112; Peptidase_C1; 1.
CC PRINTS; PR00705; PAPAIN.
CC ProDom; PD000158; Peptidase_C1; 1.
CC PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
CC PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
CC PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
KW Hydrolase; Thiol protease; Zymogen; Signal.
FT SIGNAL 1 24 POTENTIAL.
FT PROPEP 25 122
FT CHAIN 123 345 ANANAIN.
FT ACT_SITE 147 147
FT ACT_SITE 279 279 BY SIMILARITY.
FT DISULFID 144 184 BY SIMILARITY.

```

```

FT DISULFID 178 217 BY SIMILARITY.
FT DISULFID 273 325 BY SIMILARITY.
FT CONFLICT 291 324 S -> A (IN REF. 2).
FT CONFLICT 324 334 L -> I (IN REF. 2).
SQ SEQUENCE 345 AA; 38248 MW; FAF298080174D87 CRC64;

Query Match          86.1%; Score 93; DB 1; Length 345;
Best Local Similarity 94.7%; Pred. No. 1.1e-07;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKQN 19
    |||||
Db 123 VPOSIDWRDYGAVTSVKQN 141
    |||||

RESULT 3
RD19 ARATH
ID RD19 ARATH STANDARD; PRT; 368 AA.
AC P43296;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cysteine proteinase RD19A precursor (EC 3.4.22.-).
GN RD19A OR AT4G39090 OR F19H22.190.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=93314960; PubMed=8325504;
RA Koizumi M., Yamaguchi-Shinozaki K., Tsuji H., Shinozaki K.;
RT "Structure and expression of two genes that encode distinct drought-
inducible cysteine proteinases in Arabidopsis thaliana.";
RL Gene 129:175-182(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.P.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
Harris B., Anstorge W., Brandt P., Grivell L.A., Rieger M.,
Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
Kreiss M., Delsen M., Puigdomenech P., Watson M., Schmidheini T.,
Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
Langham S.-A., McCullagh B., Bilham L., Robben J.,
Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
Weitzenecker T., Bothe G., Ransperger U., Hilbert H., Braun M.,
Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
De Keyser A., Buyssebaert C., Gielens J., Villarroel R., De Clercq R.,
Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
Clark L., Doggett J., Hall S., Kay M., Lemard N., McIay K., Mayes R.,
Petrett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
Gabel C., Fuchs M., Fartmann B., Grandjean K., Dauner D., Herzl A.,
Neumann S., Argitlou F., Vitale D., Liguori R., Piravandi E.,
Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
Chedor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
Gibbons T., Weber N., Vandenbol M., Baigues M., Terol J., Torres A.,
Perez-Perez A., Fumelle B., Bent E., Johnson S., Racon D., Jesse T.,
Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,

```

RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
 RA Lareille P., Coutney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., All J., Berghoff A., Jones K., Dione K., Cotton M., Joshi C.,
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
 RA Ma P., Zhong J., Preston R., Vill D., Shekhar M., Mateo A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
 RA Granat S., Shoddy N., Hasegawa A., Hamed A., Lochi M., Johnson A.,
 RA Chen E., Marra M., Martensen R., McCombie W.R.,
 RA "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 thaliana".
 RT Nature 402:769-777(1999).
 CC -1- INDUCTION: BY HIGH SALT CONDITIONS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, D13042, BAA02373.1, -.
 DR EMBL, AL035679, CAB3829.1, -.
 DR EMBL, AL161594, CAB80572.1, -.
 DR HSSP, P07711, ICUL.
 DR MEROPS, C01.022, -.
 DR InterPro: IPR000168; Peptidase_C1.
 DR InterPro: IPR000169; SHPOT_acsite.
 DR Pfam: PF00112; Peptidase_C1.1.
 DR PRINTS: PR00705; PAPAIN_1.
 DR ProDom: PD000158; Peptidase_C1.1.
 DR PROSITE, PS00138; THIOL_PROTEASE_CYS; 1.
 DR PROSITE, PS00639; THIOL_PROTEASE_HIS; 1.
 DR PROSITE, PS00640; THIOL_PROTEASE_ASN; 1.
 KW Hydroxylase; Thiol protease; Zymogen; Glycoprotein; Signal.
 FT SIGNAL 1 23
 FT PROPEP 24 134
 FT CHAIN 135 368
 FT ACT_SITE 159 159
 FT ACT_SITE 302 302
 FT ACT_SITE 329 329
 FT DISULFID 156 206
 FT DISULFID 194 240
 FT DISULFID 296 350
 FT CARBOHYD 253 253
 SQ SEQUENCE 368 AA; 40418 MW; 9D64CF35E07F519D CRC64;
 Query Match 72.2%; Score 78; DB 1; Length 368;
 Best Local Similarity 68.4%; Pred. No. 3.2e-05;
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 VPQSIDRDYGVATVSKNO 19
 Db 135 LPEDFMDRHDGAVTVKNO 153
 RESULT 4
 ID PAPA2 CARPA STANDARD; PRT; 352 AA.
 AC P14080;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Chymopapain precursor (EC 3.4.22.6) (Papaya proteinase II) (PPII).
 OS Carica papaya (Papaya).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Caricaceae; Carica.
 OC NCBI_TaxID=3649;

RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Leaf;
 RA Connerton I.F.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 135-352.
 RX MEDLINE=90179730; PubMed=2106878;
 RA Watson D.C., Yaguchi M., Lynn K.R.;
 RT "The amino acid sequence of chymopapain from Carica papaya.";
 RL Biochem. J. 266:75-81(1990).
 RN [3]
 RP SEQUENCE OF 135-352.
 RX MEDLINE=89302685; PubMed=2500950;
 RA Jacquet A., Kleinschmidt T., Schneek A.G., Looze Y., Braunitzer G.;
 RT "The thiol proteinases from the latex of Carica papaya L. III. The
 RL primary structure of chymopapain.";
 RN Biol. Chem. Hoppe-Seyler 370:425-434(1989).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
 RX MEDLINE=97128645; PubMed=8973203;
 RA Maes D., Bouckaert J., Poortmans F., Wyns L., Looze Y.;
 RT "Structure of chymopapain at 1.7-A resolution.";
 RL Biochemistry 35:16292-16298(1996).
 CC -1- CATALYTIC ACTIVITY: Specificity similar to that of papain.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, X97789, CAA66378.1, -.
 DR PIR, S04222, S04222.
 DR PIR, S08285, S08285.
 DR PDB, 1YAL; 23-DEC-96.
 DR MEROPS, C01.002, -.
 DR InterPro: IPR000668; Peptidase_C1.
 DR InterPro: IPR000169; SHPOT_acsite.
 DR Pfam: PF00112; Peptidase_C1.1.
 DR PRINTS: PR00705; PAPAIN_1.
 DR ProDom: PD000158; Peptidase_C1.1.
 DR PROSITE, PS00138; THIOL_PROTEASE_CYS; 1.
 DR PROSITE, PS00639; THIOL_PROTEASE_HIS; 1.
 DR PROSITE, PS00640; THIOL_PROTEASE_ASN; 1.
 KW Hydroxylase; Thiol protease; Zymogen; Signal; 3D-structure.
 FT SIGNAL 1 18
 FT PROPEP 19 134
 FT CHAIN 135 352
 FT ACT_SITE 159 159
 FT ACT_SITE 293 293
 FT ACT_SITE 313 313
 FT DISULFID 156 197
 FT DISULFID 190 229
 FT DISULFID 287 338
 SQ SEQUENCE 352 AA; 39414 MW; 50EA31EBFCFA9F9F CRC64;
 Query Match 70.4%; Score 76; DB 1; Length 352;
 Best Local Similarity 83.3%; Pred. No. 6.4e-05;
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 POSIDRDYGVATVSKNO 19
 Db 136 POSIDRAKAGAVTVKNO 153
 RESULT 5
 ID CATL_PIG STANDARD; PRT; 334 AA.
 AC Q28944;

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cathepsin L precursor (EC 3.4.22.15).
 GN CTSL.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OC NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Epididymis;
 RX MEDLINE=96069841; PubMed=7492581;
 RA Okamura N., Tamba M., Uchiyama Y., Sugita Y., Dacheux F.,
 RA Syntin P., Dacheux J.L.;
 RT "Direct evidence for the elevated synthesis and secretion of
 RT procathepsin L in the distal caput epididymis of Boar.";
 RL Biochim. Biophys. Acta 1245:221-226 (1995).
 CC -!- FUNCTION: IMPORTANT FOR THE OVERALL DEGRADATION OF PROTEINS IN
 CC LYOSOMES.
 CC -!- CATALYTIC ACTIVITY: Specificity close to that of papain. As
 CC compared to cathepsin B, cathepsin L exhibits higher activity
 CC towards protein substrates, but has little activity on Z-Arg-Arg-
 CC NHMeC, and no peptidyl-di-peptidase activity.
 CC -!- SUBUNIT: DIMER OF A HEAVY AND A LIGHT CHAIN LINKED BY DISULFIDE
 CC BONDS.
 CC -!- SUBCELLULAR LOCATION: Lysosomal.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D37917; BAA07140.1; -
 DR HSSP: O60911; LFH0.
 DR MEROPS: C01.032; -
 DR InterPro: IPR000668; Peptidase_C1.
 DR Pfam: PF00112; Peptidase_C1; 1.
 DR PRINTS: PR00705; PAPAIN.
 DR ProDom: PD000158; Peptidase C1; 1.
 DR PROSITE: PS00139; THIOLESTERASE CYS; 1.
 DR PROSITE: PS00639; THIOLESTERASE HIS; 1.
 DR PROSITE: PS00640; THIOLESTERASE ASN; 1.
 DR Hydrolase; Thiol protease; Lysosome; Zymogen; Signal.
 KW SIGNAL 1 17
 FT PROPEP 18 117 ACTIVATION PEPTIDE.
 FT CHAIN 118 289 CATHEPSIN L HEAVY CHAIN.
 FT PROPEP 290 291 BY SIMILARITY.
 FT CHAIN 292 334 CATHEPSIN L LIGHT CHAIN.
 FT ACT_SITE 138 138 BY SIMILARITY.
 FT ACT_SITE 277 277 BY SIMILARITY.
 FT ACT_SITE 301 301 BY SIMILARITY.
 FT DISULFID 135 178 BY SIMILARITY.
 FT DISULFID 169 212 BY SIMILARITY.
 FT DISULFID 270 323 INTERCHAIN (BY SIMILARITY).
 FT CARBOHYD 222 222 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 292 292 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 334 AA; 37178 MW; 51DBA79ACCF2CE53 CRC64;
 Query Match 69.4%; Score 75; DB 1; Length 334;
 Best Local Similarity 68.4%; Pred. No. 8.9e-05;
 Matches 13; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 QY 1 VPOSIDWRDYGAVTSVKQ 19
 Db 114 VPKSVDRKGYVTAVKQ 132

RESULT 6
 CC3_CARCN STANDARD; PRT; 43 AA.
 AC P32556;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cysteine proteinase III (EC 3.4.22.-) (CC-III) (Fragment).
 OS Carica candamarcensis.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Caricaceae; Carica.
 OC NCBI_TaxID=29731;
 RN [1]
 RP SEQUENCE
 RC TISSUE=Latex;
 RX MEDLINE=94030669; PubMed=8216902;
 RA Walreavens V., Jaziri M., van Beeumen J., Schnek A.G.,
 RA Kleinschmidt T., Looze Y.;
 RT "Isolation and preliminary characterization of the cysteine-
 RT proteinases from the latex of Carica candamarcensis Hook.";
 RL Biol. Chem. Hoppe-Seyler 374:501-506 (1993).
 CC -!- PTM: GLYCOSYLATED.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
 DR HSSP: P14080; IVAL.
 DR MEROPS: C01.020; -
 DR InterPro: IPR000668; Peptidase_C1.
 DR InterPro: IPR000189; SHprot_acsite.
 DR Pfam: PF00112; Peptidase_C1; 1.
 DR ProDom: PD000158; Peptidase C1; 1.
 DR PROSITE: PS00639; THIOLESTERASE HIS; PARTIAL.
 DR PROSITE: PS00640; THIOLESTERASE ASN; PARTIAL.
 DR PROSITE: PS00139; THIOLESTERASE CYS; 1.
 DR Hydrolase; Thiol protease; Glycoprotein.
 FT ACT_SITE 25 25 BY SIMILARITY.
 FT NON_TER 43 43
 SQ SEQUENCE 43 AA; 4636 MW; F4C5D2881886E291 CRC64;
 Query Match 68.5%; Score 74; DB 1; Length 43;
 Best Local Similarity 77.8%; Pred. No. 1.5e-05;
 Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 POSIDWRDYGAVTSVKQ 19
 Db 2 PESIDWRKKGAVTPVKQ 19

RESULT 7
 CC4_CARCN STANDARD; PRT; 43 AA.
 AC P32557;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cysteine proteinase IV (EC 3.4.22.-) (CC-IV) (Fragment).
 OS Carica candamarcensis.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Caricaceae; Carica.
 OC NCBI_TaxID=29731;
 RN [1]
 RP SEQUENCE
 RC TISSUE=Latex;
 RX MEDLINE=94030669; PubMed=8216902;
 RA Walreavens V., Jaziri M., van Beeumen J., Schnek A.G.,
 RA Kleinschmidt T., Looze Y.;
 RT "Isolation and preliminary characterization of the cysteine-
 RT proteinases from the latex of Carica candamarcensis Hook.";
 RL Biol. Chem. Hoppe-Seyler 374:501-506 (1993).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
 DR HSSP: P14080; IVAL.
 DR MEROPS: C01.020; -
 DR InterPro: IPR000668; Peptidase_C1.

DR InterPro; IPR000169; SHProt_acsite.
 DR Pfam; PF00112; Peptidase_C1.1.
 DR ProDom; PD000158; Peptidase_C1.1.
 DR PROSITE; PS00639; THIOI_PROTEASE_HIS; PARTIAL.
 DR PROSITE; PS00640; THIOI_PROTEASE_ASN; PARTIAL.
 DR PROSITE; PS00139; THIOI_PROTEASE_CYS.1.
 KM Hydrolase; Thiol protease.
 FT ACT SITE 25 25
 FT NON_TER 43 43 BY SIMILARITY.
 SQ SEQUENCE 43 AA; 4685 MW; F4DS945A9386E291 CRC64;
 Query Match 68.5%; Score 74; DB 1; Length 43;
 Best Local Similarity 77.8%; Pred. No. 1.5e-05;
 Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 POSIDMRDYGAVTSVKNO 19
 DB 2 PESIDMRKKGAVTPVKNQ 19
 RESULT 8
 CATL_MOUSE STANDARD; PRT; 334 AA.
 AC P06797;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cathepsin L precursor (EC 3.4.22.15) (Major excreted protein) (MEP).
 GN CTSL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=88076849; Pubmed=3689328;
 RA Troen B.R., Gal S., Gottesman M.M.;
 RT "Sequence and expression of the cDNA for MEP (major excreted protein), a transformation-regulated secreted cathepsin.";
 RL Biochem. J. 246:731-735(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8821715; Pubmed=2835398;
 RA Joseph L.J., Chang L.C., Stamenkovich D., Sukhatme V.P.;
 RT "Complete nucleotide and deduced amino acid sequences of human and murine preprocathepsin L. An abundant transcript induced by transformation of fibroblasts.";
 RL J. Clin. Invest. 81:1621-1629(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8703683; Pubmed=3533924;
 RA Portnoy D.A., Erickson A.H., Kochan J., Ravetch J.V., Unkeless J.C.;
 RT "Cloning and characterization of a mouse cysteine proteinase.";
 RL J. Biol. Chem. 261:14697-14703(1986).
 RN [4]
 RP SEQUENCE FROM N.A., AND CARBOHYDRATE-LINKAGE SITE.
 RC TISSUE=Liver;
 RX MEDLINE=9112761; Pubmed=2275556;
 RA Stearns N.A., Dong J., Pan J.X., Brenner D.A., Sahagian G.G.;
 RT "Comparison of cathepsin L synthesized by normal and transformed cells at the gene, message, protein, and oligosaccharide levels.";
 RL Arch. Biochem. Biophys. 283:447-457(1990).
 RN [5]
 RP SEQUENCE OF 89-300 FROM N.A.
 RC STRAIN=BNL;
 RX MEDLINE=86271744; Pubmed=3755373;
 RA Denhardt D.T., Hamilton R.T., Parfett C.L.D., Edwards D.R.,
 RA Pierre R.S., Waterhouse P., Nilsson-Hamilton M.;
 RT "Close relationship of the major excreted protein of transformed murine fibroblasts to thiol-dependent cathepsins.";
 RL Cancer Res. 46:4590-4593(1986).
 CC -1- FUNCTION: IMPORTANT FOR THE OVERALL DEGRADATION OF PROTEINS IN LYSOSOMES.

CC -1- CATALYTIC ACTIVITY: Specificity close to that of papain. As compared to cathepsin B, cathepsin L exhibits higher activity towards protein substrates, but has little activity on Z-Arg-Arg-NMec, and no peptidyl-dipeptidase activity.
 CC -1- SUBUNIT: DIMER OF A HEAVY AND A LIGHT CHAIN LINKED BY DISULFIDE BONDS.
 CC -1- SUBCELLULAR LOCATION: Lysosomal.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X06086; CA929470.1; -
 DR EMBL; U02583; AA937445.1; -
 DR EMBL; M20495; AA93984.1; -
 DR EMBL; X04392; CA927980.1; -
 DR PIR; S01177; KHM5L.
 DR PIR; S13890; S13890.
 DR HSSP; P07711; 1CJL.
 DR MEROPS; C01.032; -.
 DR MGD; MGI:88564; Ctsl.
 DR InterPro; IPR000668; Peptidase_C1.
 DR InterPro; IPR000169; SHProt_acsite.
 DR Pfam; PF00112; Peptidase_C1.1.
 DR PRINTS; PR00705; PAFAIN.
 DR ProDom; PD000158; Peptidase_C1.1.
 DR PROSITE; PS00139; THIOI_PROTEASE_CYS.1.
 DR PROSITE; PS00639; THIOI_PROTEASE_HIS.1.
 DR PROSITE; PS00640; THIOI_PROTEASE_ASN.1.
 KM Hydrolase; Thiol protease; Glycoprotein; Lysosome; Zymogen; Signal.
 FT SIGNAL 1 17
 FT PROPEP 18 113
 FT TROEN 114 288
 FT CHAIN 291 334
 FT ACT_SITE 138 138
 FT ACT_SITE 276 276
 FT ACT_SITE 300 300
 FT DISULFID 135 178
 FT DISULFID 169 211
 FT DISULFID 269 322
 FT CARBOHYD 221 221
 FT CONFLICT 58 58
 FT CONFLICT 177 177
 SQ SEQUENCE 334 AA; 37547 MW; FE674704307AD98 CRC64;
 Query Match 68.5%; Score 74; DB 1; Length 334;
 Best Local Similarity 63.2%; Pred. No. 0.00013;
 Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
 QY 1 VPOSIDMRDYGAVTSVKNO 19
 DB 114 IPKSYDMREKGCVTVPVKNQ 132
 RESULT 9
 A494_ARATH STANDARD; PRT; 361 AA.
 ID A494_ARATH
 AC P43295; Q9SUT5;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable cysteine proteinase A494 precursor (EC 3.4.22.-).
 GN AT2G21430 OR F3K23.19.
 GN Arabidopsis thaliana (Mouse-ear cress).
 OS Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;

```

DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cathepsin L precursor (EC 3.4.22.15) (SMCL1).
GN CL1.
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OC NCBI_TaxID=6183;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=Puerto Rican;
RX MEDLINE=95140043; PubMed=7838171;
RA Smith A.M., Dalton J.P., Clough K.A., Kilbane C.L., Harrop S.A.,
RA Hole N., Brindley P.J.;
RT "Adult Schistosoma mansoni express cathepsin L proteinase activity.";
RL Mol. Biochem. Parasitol. 67:11-19(1994)
CC -!- FUNCTION: MAY BE CRUCIAL FOR METABOLISM OF HOST HEMOGLOBIN.
CC -!- CATALYTIC ACTIVITY: Specificity close to that of papain. As
CC compared to cathepsin B, cathepsin L exhibits higher activity
CC towards protein substrates, but has little activity on Z-Arg-Arg-
CC NHec, and no peptidyl-dipeptidase activity.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL: U07345; AAC6485.1; --
CC HSPF: O60911; LFPO.
DR MEROPS; COI.018; --
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR000169; SHprot_accsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPAIN.
DR ProDom; PD000158; Peptidase_C1; 1.
DR PROSITE; PS00139; THIOL PROTEASE CYS; 1.
DR PROSITE; PS00639; THIOL PROTEASE HIS; 1.
DR PROSITE; PS00640; THIOL PROTEASE ASN; 1.
DR Hydrolase; Thiol protease; Glycoprotein; Lysosome; Zymogen; Signal.
KW SIGNAL
FT SIGNAL 1 ?
FT PROPEL 2 104 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 105 319 CATHEPSIN L.
FT ACT_SITE 129 129 BY SIMILARITY.
FT ACT_SITE 265 265 BY SIMILARITY.
FT ACT_SITE 286 286 BY SIMILARITY.
FT DISULFID 126 167 BY SIMILARITY.
FT DISULFID 160 200 BY SIMILARITY.
FT DISULFID 258 307 BY SIMILARITY.
FT SEQUENCE 319 AA; 36136 MW; 9CFE68A3F1193479 CRC64;
Query Match 67.6%; Score 73; DB 1; Length 319;
Best Local Similarity 63.2%; Pfam NO. 0.00018;
Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 1 VPSGIDWRDYGAVTSVKNQ 19
DB 105 IPKNFDRNREKGAIVEVKNQ 123
:|||||:|||||
RESULT 11
ID CAT3_FASHE STANDARD; PRT; 19 AA.
AC P80532;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative cathepsin L3 (EC 3.4.22.15) (Newly excysed juvenile protein

```


DE 8) (Fragment).
 OS Fasciola hepatica (Liver fluke).
 CC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 OC Echinostomidae; Echinostomata; Fasciolidae; Fasciola.
 NC NCB1_TaxID=6192;
 RN [1]
 RP SEQUENCE.
 RA MEDLINE=9536693; PubMed=7639732;
 RX Tkalcic J., Ashman K., Meusen E.;
 RT "Fasciola hepatica: rapid identification of newly excysted juvenile
 proteins";
 RL Biochem. Biophys. Res. Commun. 213:169-174(1995).
 CC -1- FUNCTION: THIOL PROTEASE.
 CC -1- CATALYTIC ACTIVITY: Specificity close to that of papain. As
 compared to cathepsin B, cathepsin L exhibits higher activity
 toward protein substrates, but has little activity on Z-Arg-Arg-
 NHec, and no peptidyl-diesterase activity.
 CC -1- SUBUNIT: DIMER OF AN HEAVY AND A LIGHT CHAIN LINKED BY DISULFIDE
 BONDS (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Lysosomal (potential).
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT THE NEWLY ECYSTED JUVENILE
 STAGE.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
 CC MEROPS; C01.033; -;
 DR Interpro: IPR000668; Peptidase_C1.
 DR Interpro: IPR000169; SHprot_acsite.
 DR Pfam: PF00112; Peptidase_C1; 1.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; PARTIAL.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; PARTIAL.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASN; PARTIAL.
 DR Hydrolase; Thiol protease; Lysosome.
 FT NON TER 19
 FT SEQUENCE 19 AA; 2242 MW; 53FFB5835BECB0D7 CRC64;
 SQ
 Query Match 66.7%; Score 72; DB 1; Length 19;
 Best Local Similarity 72.2%; Pred. No. 1.4e-05;
 Matches 13; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPOSIDMRDYGAVTSVKV 18
 |||||:|||||
 DB 2 VPASIDMRDYGAVTSVKD 19
 :|||:|||||

RESULT 12
 ID ORYB_ORYSA STANDARD; PRT; 471 AA.
 AC P25777;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Oryza sativa chain precursor (EC 3.4.22.-).
 OS Oryza sativa (Rice).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC Eriocarpaceae; Oryzaceae; Oryza.
 NC NCB1_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare; TISSUE=Seed;
 RX MEDLINE=91358494; PubMed=1885617;
 RA Watanabe H., Abe K., Emori Y., Hosoyama H., Arai S.;
 RT "Molecular cloning and gibberellin-induced expression of multiple
 RT cysteine proteinases of rice seeds (oryza sativa).";
 RL J. Biol. Chem. 266:16897-16902(1991).
 CC -1- TISSUE SPECIFICITY: EXPRESSED ONLY IN SEEDS.
 CC -1- INDUCTION: BY GIBBERELLIC ACID (GA).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D90407; BA14403.1; -;
 DR FIR; J00389; KHRZOB.
 DR HSP; P00785; 2ACT.
 DR MEROPS; C01.029; -;
 DR Interpro: IPR000118; Granulin.
 DR Interpro: IPR000668; Peptidase_C1.
 DR Interpro: IPR000169; SHprot_acsite.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR Pfam; PF00396; Granulin; 1.
 DR PRINTS; PR00705; PAPA1N.
 DR Prodom; P0000158; peptidase_C1; 1.
 DR SMART; SMO0277; GRAN; 1.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
 DR Hydrolase; Thiol protease; Zymogen; Glycoprotein; signal.
 FT SIGNAL 1 21
 FT PROPEP 22 139
 FT CHAIN 140 471
 FT ACT_SITE 164 471
 FT ACT_SITE 164 471
 FT ACT_SITE 301 301
 FT ACT_SITE 321 321
 FT ACT_SITE 321 321
 FT DISULFID 161 204
 FT DISULFID 195 237
 FT DISULFID 295 346
 FT CARBOHYD 340 340
 FT CARBOHYD 388 388
 FT SEQUENCE 471 AA; 50505 MW; 0BF39D33995CEB1D CRC64;
 SQ
 Query Match 66.7%; Score 72; DB 1; Length 471;
 Best Local Similarity 63.2%; Pred. No. 0.00039;
 Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPOSIDMRDYGAVTSVKV 19
 :|||:|||||
 DB 140 LPESVDMREKAVAPVKNQ 158
 :|||:|||||

RESULT 13
 ID CC2_CARCN STANDARD; PRT; 43 AA.
 AC P32955;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cysteine proteinase II (EC 3.4.22.-) (CC-II) (Fragment).
 OS Carica candamarcensis.
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC Eucosids II; Brassicales; Caricaceae; Carica.
 NC NCB1_TaxID=29731;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Latex;
 RX MEDLINE=94030669; PubMed=8216902;
 RA Walteavens V., Jaziri M., van Beunemen J., Schneek A.G.,
 RA Kleinschmidt T., Looze Y.;
 RT "Isolation and preliminary characterization of the cysteine-
 RT proteinases from the latex of Carica candamarcensis Hook.";
 RL J. Biol. Chem. Hoppe-Sevler 374:501-506(1993).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
 CC HSP; P14080; 1VAL.
 DR MEROPS; C01.09A; -;
 DR Interpro: IPR000668; Peptidase_C1.
 DR Interpro: IPR000169; SHprot_acsite.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR Prodom; P0000158; Peptidase_C1; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; PARTIAL.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASN; PARTIAL.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.

KW Hydrolase; Thiol protease.
FT ACT SITE 25 25 BY SIMILARITY.
FT NON_TER 43 43
SQ SEQUENCE 43 AA; 4568 MW; 1777ACB74BC5F34B CRC64;

Query Match 65.7%; Score 71; DB 1; Length 43;
Best Local Similarity 68.4%; Pred. No. 4.7e-05;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 PQSIDWRDYGAVTSVKNQ 20
| : | | | | : | | | | : | | | |
Db 2 PGSDVDRQKGA VTPVKDQN 20

RESULT 14

GPII_ZINOF STANDARD; PRT; 221 AA.
AC P82474;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cysteine proteinase GP-II (EC 3.4.22.-).
OS Zingiber officinale (Ginger).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Zingiberales; Zingiberaceae;
OC Zingiber.
OX NCBI_TaxID=94328;
RN [1]_TaxID=94328;
RP SEQUENCE.

RC TISSUE=Root;
RX MEDLINE=20156257; PubMed=10691991;
RA Choi K.H.; Laursen R.A.;
RT "Amino-acid sequence and glycan structures of cysteine proteases with
RT proline specificity from ginger rhizome Zingiber officinale.";
RL Eur. J. Biochem. 267:1516-1526(2000).
CC -|- CATALYTIC ACTIVITY: Preferential cleavage with a proline residue
at p2.

CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.

DR HSSP; P00785; 2ACT.
DR GlycoSuiteDB; P82474; -.
DR InterPro; IPR000668; Peptidase C1.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPA1N.
DR ProDom; PD000158; Peptidase C1; 1.
DR PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.
DR PROSITE; PS00640; THIOI_PROTEASE_ASN; FALSE_NEG.
KW Hydrolase; Thiol protease; Glycoprotein.
FT ACT_SITE 27, 27 BY SIMILARITY.
FT ACT_SITE 161, 161 BY SIMILARITY.
FT DISULFID 24, 65 BY SIMILARITY.
FT DISULFID 58, 98 BY SIMILARITY.
FT DISULFID 155, 206 BY SIMILARITY.
FT CARBOHYD 99, 99 N-LINKED (GLCNAC. . .).
FT CARBOHYD 156, 156 /FTID=CAR_000190.
FT CARBOHYD 156, 156 N-LINKED (GLCNAC. . .).
FT /FTID=CAR_000200.
SQ SEQUENCE 221 AA; 23922 MW; 909A312BD8632D42 CRC64;

Query Match 65.7%; Score 71; DB 1; Length 221;
Best Local Similarity 68.4%; Pred. No. 0.00026;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 VQSIDWRDYGAVTSVKNQ 19
: | | | | : | | | | : | | | |
Db 3 LPDSIDWREKGA VVPVKNO 21

RESULT 15

GPI_ZINOF STANDARD; PRT; 221 AA.
ID_GPI_ZINOF
AC P82473;

DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cysteine proteinase GP-I (EC 3.4.22.-).
OS Zingiber officinale (Ginger).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Zingiberales; Zingiberaceae;
OC Zingiber.
OX NCBI_TaxID=94328;
RN [1]
RP SEQUENCE.
RC TISSUE=Root;
RX MEDLINE=20156257; PubMed=10691991;
RA Choi K.H.; Laursen R.A.;
RT "Amino-acid sequence and glycan structures of cysteine proteases with
RT proline specificity from ginger rhizome Zingiber officinale.";
RL Eur. J. Biochem. 267:1516-1526(2000).
CC -|- CATALYTIC ACTIVITY: Preferential cleavage with a proline residue
at p2.

CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.

CC -|- CAUTION: THE AUTHORS REGARD THE SEQUENCE AS TENTATIVE, AS THEY

CC BELIEVE THAT IT MAY HAVE BEEN CONTAMINATED BY A HOMOLOGOUS

CC PROTEIN.

DR HSSP; P00785; 2ACT.

DR InterPro; IPR000668; Peptidase C1.

DR Pfam; PF00112; Peptidase_C1; 1.

DR PRINTS; PR00705; PAPA1N.

DR ProDom; PD000158; Peptidase C1; 1.

DR PROSITE; PS00139; THIOI_PROTEASE_CYS; FALSE_NEG.

DR PROSITE; PS00639; THIOI_PROTEASE_HIS; FALSE_NEG.

DR PROSITE; PS00640; THIOI_PROTEASE_ASN; 1.

KW Hydrolase; Thiol protease; Glycoprotein.

FT ACT_SITE 27, 27 BY SIMILARITY.

FT ACT_SITE 161, 161 BY SIMILARITY.

FT DISULFID 24, 65 BY SIMILARITY.

FT DISULFID 58, 98 BY SIMILARITY.

FT DISULFID 155, 206 BY SIMILARITY.

FT CARBOHYD 95, 95 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 156, 156 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT UNSURE 2, 2 OR D.

SQ SEQUENCE 221 AA; 24241 MW; 3035D7870EA743DB CRC64;

Query Match 65.7%; Score 71; DB 1; Length 221;

Best Local Similarity 68.4%; Pred. No. 0.00026;

Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 VQSIDWRDYGAVTSVKNO 19

: | | | | : | | | | : | | | |

Db 3 LPDSIDWREKGA VVPVKNO 21

Search completed: November 21, 2002, 16:09:01

Job time : 8.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2002, 16:06:44 ; Search time 25 Seconds

(without alignments)
164,838 Million cell updates/sec

Title: US-09-674-738-2
Perfect score: 105
Sequence: 1 VPOSIDWRDSCGAVTSVKNKG 20

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mbc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	105	100.0	20	10 Q9S7J0	Q9S7J0 ananas como
2	94	89.5	20	10 Q9S8M1	Q9S8M1 ananas como
3	93	88.6	20	10 Q9S965	Q9S965 ananas como
4	93	88.6	356	10 Q81084	Q81084 ananas como
5	93	88.6	356	10 Q23799	Q23799 ananas como
6	88	88.8	454	10 Q40922	Q40922 pseudotsuga
7	85	81.0	334	5 Q9US54	Q9US54 tetrahy aura
8	84	80.0	226	10 Q9SMH9	Q9SMH9 carica papa
9	84	80.0	227	10 Q9SMH8	Q9SMH8 carica papa
10	84	80.0	324	10 Q23801	Q23801 ananas como
11	84	80.0	337	10 Q8S334	Q8S334 lycopersico
12	84	80.0	351	10 Q23791	Q23791 ananas como
13	84	80.0	352	10 Q9SMI1	Q9SMI1 carica papa
14	84	80.0	361	10 Q9SMI0	Q9SMI0 carica papa
15	83	79.0	357	10 Q81085	Q81085 ananas como
16	83	79.0	367	10 Q40261	Q40261 mesembryant

17	83	79.0	374	5 Q18455	Q18455 heterodera
18	83	79.0	381	5 Q9CQK6	Q9CQK6 leishmania
19	83	79.0	394	5 Q95WR6	Q95WR6 leishmania
20	83	79.0	443	5 P90628	P90628 leishmania
21	83	79.0	443	5 Q95WR7	Q95WR7 leishmania
22	83	79.0	443	5 Q18698	Q18698 leishmania
23	83	79.0	443	5 P92185	P92185 leishmania
24	83	79.0	489	5 Q14544	Q14544 caenorhabdi
25	82	78.1	189	5 Q966T7	Q966T7 spiometra
26	82	78.1	334	11 Q9D0C0	Q9D0C0 mus musculus
27	82	78.1	334	11 Q21U20	Q21U20 mus musculus
28	82	78.1	336	5 Q95NF1	Q95NF1 spiometra
29	82	78.1	336	5 Q02586	Q02586 spiometra
30	82	78.1	339	5 Q26636	Q26636 sarcophaga
31	82	78.1	359	10 Q41697	Q41697 vicia sativ
32	82	78.1	360	10 Q6S039	Q6S039 ricinus com
33	81	77.1	364	10 Q9STL5	Q9STL5 arabidopsis
34	80	76.2	224	5 Q11003	Q11003 schistosoma
35	80	76.2	315	10 Q9LM66	Q9LM66 arabidopsis
36	80	76.2	326	5 Q9NMH3	Q9NMH3 fasciola gi
37	80	76.2	326	5 Q8T0W9	Q8T0W9 fasciola he
38	80	76.2	331	5 Q46031	Q46031 sitophilus
39	80	76.2	331	5 Q46032	Q46032 sitophilus
40	80	76.2	333	6 Q9GL24	Q9GL24 canis famli
41	80	76.2	337	5 Q9SV59	Q9SV59 delia radic
42	80	76.2	338	5 Q46030	Q46030 sitophilus
43	80	76.2	340	5 Q9NH99	Q9NH99 stylyonchia
44	80	76.2	356	10 Q9SYO2	Q9SYO2 arabidopsis
45	80	76.2	360	10 Q82708	Q82708 pisum sativ

ALIGNMENTS

RESULT 1

Q9S7J0 PRELIMINARY; PRT; 20 AA.

AC Q9S7J0
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Bromelain (Fragment).
OS Ananas comosus (Pineapple).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae; Ananas.
OC NCBI_TaxID=4615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95298192; PubMed=779262;
RA Harrach T., Eckert K., Schulze-Forster K., Nuck R., Grunow D.,
RA Maurer H.R.;
RT "Isolation and partial characterization of basic proteinases from stem
bromelain.";
RN [2]
RP J. Protein Chem. 14:41-52(1995).
RX SEQUENCE.
RX MEDLINE=9430946; PubMed=8033898;
RA Napper A.D., Bennett S.P., Borowski M., Holdridge M.B., Leonard M.J.,
RA Rogers E.E., Duan Y., Laursen R.A., Reinhold B., Shames S.L.;
RL Biochem. J. 301:727-735(1994).
DR InterPro; IPR000668; Peptidase C1.
PF pfam: PF00112; Peptidase C1; 1.
SQ SEQUENCE 20 AA; 2144 MW; 83A842BC812C3EBC CRC64;

Query Match 100.0%; Score 105; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.3e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VPOSIDWRDSCGAVTSVKNKG 20
DB 1 VPOSIDWRDSCGAVTSVKNKG 20

```

RESULT 2
Q9S8M1 ID Q9S8M1 PRELIMINARY; PRT; 20 AA.
AC Q9S8M1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE COMOSAIN (Fragment)
OS Ananas comosus (Pineapple)
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae incertae sedis;
OC Bromeliaceae; Ananas.
OX NCBI_TaxID=4615;
RN [1]
RP SEQUENCE 20 AA; 2219 MW; 83A94A16232C3EBC CRC64;
RX MEDLINE=94330946; PubMed=8053898;
RA Napper A.D., Bennett S.P., Borowski M., Holdridge M.B., Leonard M.J.,
RA Rogers E.E., Duan Y., Laursen R.A., Reinhold B., Shames S.L.;
RL Biochem. J. 301:727-735(1994)
DR InterPro: IPR000668; Peptidase_C1.
DR Pfam: PF00112; Peptidase_C1; 1.
SQ SEQUENCE 20 AA; 2219 MW; 83A94A16232C3EBC CRC64;

Query Match 89.5%; Score 94; DB 10; Length 20;
Best Local Similarity 90.0%; Pred. No. 9.1e-09;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPOSIDWRDMSGAVTSVKQK 20
Db 1 VPOSIDWRNFGAVTSVKQK 20

RESULT 3
Q9S965 ID Q9S965 PRELIMINARY; PRT; 20 AA.
AC Q9S965;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Bromelain (Fragment)
OS Ananas comosus (Pineapple)
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae incertae sedis;
OC Bromeliaceae; Ananas.
OX NCBI_TaxID=4615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95298192; PubMed=7779262;
RA Harrach T., Eckert K., Schulze-Forster K., Nuck R., Grunow D.,
RA Maurer H.R.;
RT "Isolation and partial characterization of basic proteinases from stem bromelain.";
RL J. Protein Chem. 14:41-52(1995)
DR InterPro: IPR000668; Peptidase_C1.
DR Pfam: PF00112; Peptidase_C1; 1.
SQ SEQUENCE 20 AA; 2277 MW; 8F984A1E812C3EBC CRC64;

Query Match 88.6%; Score 93; DB 10; Length 20;
Best Local Similarity 94.7%; Pred. No. 1.3e-08;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPOSIDWRDMSGAVTSVKQK 19
Db 1 VPOSIDWRDYGAVTSVKQK 19

RESULT 4
Q81084 ID Q81084 PRELIMINARY; PRT; 356 AA.
AC Q81084;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)

```

```

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Cysteine proteinase, AN8 precursor.
GN AN8.
OS Ananas comosus (Pineapple).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae incertae sedis;
OC Bromeliaceae; Ananas.
OX NCBI_TaxID=4615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. SMOOTH CAYENNE; TISSUE=STEM;
RA Robertson C.E., Goodenough P.W.;
RT "Cloning and expression of an ananin gene from pineapple.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ009829; CAA08860.1; -.
DR HSSP; P07711; 1CJL.
DR MEROPS; C01.005; -.
DR InterPro: IPR000668; Peptidase_C1.
DR InterPro: IPR000169; SHprot_acSite.
DR Pfam: PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPAIN.
DR ProDom; PD000158; Peptidase_C1; 1.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW Hydrolase; Signal; Thiol protease.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 356 CYSTEINE PROTEINASE PRECURSOR, AN8.
SQ SEQUENCE 356 AA; 39528 MW; 97D705C62EC01DB8 CRC64;

Query Match 88.6%; Score 93; DB 10; Length 356;
Best Local Similarity 94.7%; Pred. No. 3.4e-07;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPOSIDWRDMSGAVTSVKQK 19
Db 124 VPOSIDWRDYGAVTSVKQK 142

RESULT 5
Q23799 ID Q23799 PRELIMINARY; PRT; 356 AA.
AC Q23799;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE F5B precursor.
OS Ananas comosus (Pineapple).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae incertae sedis;
OC Bromeliaceae; Ananas.
OX NCBI_TaxID=4615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. N67-10; TISSUE=FRUIT;
RA Muta E., Okamoto Y., Ota S.;
RT "Cloning and sequencing of cysteine proteinases in Ananas comosus.";
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; D38532; BAA22544.1; -.
DR HSSP; P07711; 1CJL.
DR MEROPS; C01.005; -.
DR InterPro: IPR001064; Crystallin.
DR InterPro: IPR000668; Peptidase_C1.
DR InterPro: IPR000169; SHprot_acSite.
DR Pfam: PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPAIN.
DR ProDom; PD000158; Peptidase_C1; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW Hydrolase; Signal; Thiol protease.
FT SIGNAL 1 24 POTENTIAL.

```

FT CHAIN 123 338 FBSB.
SQ SEQUENCE 356 AA; 39560 MW; 16C2DCCB47238928 CRC64;

Query Match
Best Local Similarity 88.6%; Score 93; DB 10; Length 356;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VPOSIDWRDGAVTAVKNG 19
DB 124 VPOSIDWRDGAVTAVKNG 142

RESULT 6
O40922 PRELIMINARY; PRT; 454 AA.

AC O40922; Q40919; TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE PSEUDOTZAIN (Cysteine protease).
PM33CASP.
OS Pseudotsuga menziesii (Douglas-fir).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Streptophyta; Coniferopsida; Coniferales; Pinaceae; Pseudotsuga.
OX NCBI_TaxID=3357;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=(MIRB) FRANCO.
RA Translanger T.J., Mista S.;
RL Physiol. Plantarum 0:0-0(0).
RN [13]
RP SEQUENCE OF 297-347 FROM N.A.
RC TISSUE=WHOLE SEEDLING;
RA Translanger T.J., Mista S.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U41902; AAC9455.1; -;
DR EMBL; 249765; CAA89835.1; -;
DR HSSP; P07711; 1CUL.
DR MEROPS; C01.029; -;
DR InterPro; IPR000118; Granulin.
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR001211; PhospholipaseA2.
DR InterPro; IPR00169; SHprot_acsite.
DR Pfam; PF00396; granulin_1.
DR PRINTS; PR00705; PAPAIV.
DR ProDom; PD000158; Peptidase_C1; 1.
DR SMART; SM00277; GRAN; 1.
DR PROSITE; PS00118; PAZ_HIS; UNKNOWN 1.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW Hydrolase; Protease; Thiol protease.
SQ SEQUENCE 454 AA; 49719 MW; 48C1938FDD6A023F CRC64;

Query Match
Best Local Similarity 83.8%; Score 88; DB 10; Length 454;
Matches 15; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 VPOSIDWRDGAVTAVKNG 20
DB 132 LPESIDWRKGAVTAVKNG 151

RESULT 7

O9U554
ID O9U554 PRELIMINARY; PRT; 334 AA.

AC O9U554;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Silicatein beta.
OS Tethya aurantia.
OC Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;
OC Hadromerida; Tethyidae; Tethya.
OX NCBI_TaxID=34494;
RN [1]
RP SEQUENCE FROM N.A.
RA Shimizu S., Lawrence C., Morse D.E.;
RL "Cloning of silicatein beta."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF098670; AAF21819.1; -;
DR HSSP; P07711; 1CUL.
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR00169; SHprot_acsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPAIV.
DR ProDom; PD000158; Peptidase_C1; 1.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW Hydrolase; Thiol protease.
SQ SEQUENCE 334 AA; 36787 MW; 4E6754BE2799065 CRC64;

Query Match
Best Local Similarity 81.0%; Score 85; DB 5; Length 334;
Matches 15; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 POSIDWRDGAVTAVKNG 20
DB 119 PESIDWRKGAVTAVKNG 137

RESULT 8
O9SMH9 PRELIMINARY; PRT; 226 AA.

AC O9SMH9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Chymopapain isoform IV (EC 3.4.22.6) (Fragment).
GN CHYMOTIV.
OS Carica papaya (papaya).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosida II; Brassicales; Caricaceae; Carica.
OX NCBI_TaxID=3649;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RA Taylor M.A., Al-Sheikh M., Revelle D.F., Sumner I.G., Connerton I.F.;
RL "cDNA cloning and expression of Carica papaya prochymopapain isoforms
RT in Escherichia coli."
RL Plant Sci. 145:41-47(1999).
DR EMBL; AJ111997; CAB38316.1; -;
DR HSSP; P14080; 1YAL.
DR MEROPS; C01.002; -;
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR00169; SHprot_acsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPAIV.
DR ProDom; PD000158; Peptidase_C1; 1.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW Hydrolase; Thiol protease.
NON_TER 1
FT SEQUENCE 226 AA; 24597 MW; 997A1B60CE49C35 CRC64;

Query Match
Best Local Similarity 80.0%; Score 84; DB 10; Length 226;

Best Local Similarity 84.2%; Pred. No. 6.5e-06;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 POSIDWRDSGAVTSVKNOG 20
DB 1 POSIDWRDSGAVTSVKNOG 19

RESULT 9
Q9SMH8 PRELIMINARY; PRT; 227 AA.
AC Q9SMH8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 20, Last annotation update)
DE Chynopapain isoform V (EC 3.4.22.6) (Fragment).
GN CHYMOV.
OS Carica papaya (Papaya).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Caricaceae; Carica.
OX NCBI_TaxID=3649;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RA Taylor M.A., Al-Sheikh M., Revell D.F., Summer I.G., Connerton I.F.;
RT "cDNA cloning and expression of Carica papaya prochymopapain isoforms
in Escherichia coli.";
RL Plant Sci. 145:41-47(1999).
DR EMBL; AJ131998; CAB38317.1; -.
DR HSSP; P14080; 1YAL.
DR MEROPS; C01.002; -.
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPAIN.
DR PRODOM; PD000158; Peptidase_C1; 1.
DR PROSITE; PS00640; THIOI_PROTEASE_ASN; 1.
DR PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.
KW Hydrolase; Thiol protease.
FT NON_TER 1
FT CHAIN 95 312
SQ SEQUENCE 227 AA; 24721 MW; 1C7C9D27AFAA80 CRC64;

Query Match 80.0%; Score 84; DB 10; Length 227;
Best Local Similarity 84.2%; Pred. No. 6.5e-06;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 POSIDWRDSGAVTSVKNOG 20
DB 2 POSIDWRDSGAVTSVKNOG 20

RESULT 10
O23801 PRELIMINARY; PRT; 324 AA.
AC O23801;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE FB1035 (Fragment).
OS Ananas comosus (Pineapple).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae incertae sedis;
OC Bromeliaceae; Ananas.
OX NCBI_TaxID=4615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. N67-10; TISSUE=FRUIT;
RA Muta E., Okamoto Y., Ota S.;
RT "Cloning and sequencing of cysteine proteinases in Ananas comosus.";
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; D38534; BAA22546.1; -.

DR HSSP; P07711; 1CJUL.
DR MEROPS; C01.028; -.
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPAIN.
DR PRODOM; PD000158; Peptidase_C1; 1.
DR PROSITE; PS00640; THIOI_PROTEASE_ASN; 1.
DR PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOI_PROTEASE_HIS; UNKNOWN_1.
KW Hydrolase; Thiol protease.
FT NON_TER 1
FT CHAIN 95 312
SQ SEQUENCE 324 AA; 36038 MW; 057462C59424339F CRC64;

Query Match 80.0%; Score 84; DB 10; Length 324;
Best Local Similarity 84.2%; Pred. No. 9.7e-06;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSGAVTSVKNOG 19
DB 96 VPOSIDWRDSGAVTSVKNOG 114

RESULT 11
Q8S334 PRELIMINARY; PRT; 337 AA.
AC Q8S334;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cysteine protease.
GN RCR3.
OS Lycopersicon pennellii (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=28526;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11976458;
RA Mulder J., Jones J.D.;
RA Kruger J., Thomas C.M., Golstein C., Dixon M.S., Smoker M., Tang S.,
RT "A Tomato Cysteine Protease Required for Cf-2-Dependent Disease
Resistance and Suppression of Autonecrosis.";
RL Science 296:744-747(2002).
DR EMBL; AF493233; AAM19208.1; -.
KW Protease.
SQ SEQUENCE 337 AA; 37389 MW; CE76884456564FBE CRC64;

Query Match 80.0%; Score 84; DB 10; Length 337;
Best Local Similarity 70.0%; Pred. No. 1e-05;
Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSGAVTSVKNOG 20
DB 123 MPSNLDWRDSGAVTSVKNOG 142

RESULT 12
O23791 PRELIMINARY; PRT; 351 AA.
AC O23791;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Bromelain precursor.
OS Ananas comosus (Pineapple).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae incertae sedis;
OC Bromeliaceae; Ananas.
OX NCBI_TaxID=4615;
RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN=CV.N67-10; TISSUE=FRUIT;
RA Muta E.; Aramaki H.; Takata Y.; Kono A.; Okamoto Y.; Oca S.;
RT "Cloning and sequencing of fruit bromelain.";
RL Submitted (JAN-1993) to the EMBL/Genbank/DBJ databases.
DR EMBL; D14059; BAA21849.1; -.
DR HSSP; P07711; ICUL.
DR MEROPS; C01.028; -.
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF00112; Peptidase_C1_1.
DR PRINTS; PR00705; PAPA1N_C1_1.
DR PRODOM; PD000158; Peptidase_C1_1.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN_1.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS_1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KM Hydrolyase; Signal; Thiol protease.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 122 339 BROMELAIN.
SQ SEQUENCE 351 AA; 39055 MW; 33781C5514424280 CRC64;

Query Match 80.0%; Score 84; DB 10; Length 351;
Best Local Similarity 84.2%; Pred. No. 1.1e-05;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 VPOSIDMRDSCAVTSYKNOG 19
Db 123 VPOSIDMRDYGAVNEVKNQ 141

RESULT 13
O9SM11 PRELIMINARY; PRT; 352 AA.
ID O9SM11;
AC O9SM11;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Chymopapain isoform II precursor (EC 3.4.22.6).
GN CHYMOTII.
OS Carica papaya (Papaya).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Caricaceae; Carica.
OX NCBI_TaxID=3649;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RA Taylor M.A.; Al-Sheikh M.; Revell D.F.; Sumner I.G.; Connerton I.F.;
RT "cDNA cloning and expression of Carica papaya prochymopapain isoforms
in Escherichia coli.";
RL Plant Sci. 145:41-47(1999).
DR EMBL; AJ131995; CAB38314.1; -.
DR HSSP; P14080; IYAL.
DR MEROPS; C01.002; -.
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF00112; Peptidase_C1_1.
DR PRINTS; PR00705; PAPA1N_C1_1.
DR PRODOM; PD000158; Peptidase_C1_1.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN_1.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS_1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KM Hydrolyase; Signal; Thiol protease.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 135 352 CHYMOPAPAIN II.
SQ SEQUENCE 352 AA; 39466 MW; A1BB21FAECF08235 CRC64;

Query Match 80.0%; Score 84; DB 10; Length 352;
Best Local Similarity 84.2%; Pred. No. 1.1e-05;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 POSIDMRDSCAVTSYKNOG 20
Db 123 VPOSIDMRDYGAVNEVKNQ 141

```

```

Db 136 POSIDMRDSCAVTSYKNOG 154

RESULT 14
O9SM10 PRELIMINARY; PRT; 361 AA.
ID O9SM10;
AC O9SM10;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Chymopapain isoform III precursor (EC 3.4.22.6).
GN CHYMOTII.
OS Carica papaya (Papaya).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Caricaceae; Carica.
OX NCBI_TaxID=3649;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RA Taylor M.A.; Al-Sheikh M.; Revell D.F.; Sumner I.G.; Connerton I.F.;
RT "cDNA cloning and expression of Carica papaya prochymopapain isoforms
in Escherichia coli.";
RL Plant Sci. 145:41-47(1999).
DR EMBL; AJ131996; CAB38315.1; -.
DR HSSP; P14080; IYAL.
DR MEROPS; C01.002; -.
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF00112; Peptidase_C1_1.
DR PRINTS; PR00705; PAPA1N.
DR PRODOM; PD000158; Peptidase_C1_1.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN_1.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS_1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KM Hydrolyase; Signal; Thiol protease.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 135 361 CHYMOPAPAIN III.
SQ SEQUENCE 361 AA; 40478 MW; EEB2A6A6BD5C0C CRC64;

Query Match 80.0%; Score 84; DB 10; Length 361;
Best Local Similarity 84.2%; Pred. No. 1.1e-05;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 POSIDMRDSCAVTSYKNOG 20
Db 136 POSIDMRDSCAVTSYKNOG 154

RESULT 15
O81085 PRELIMINARY; PRT; 357 AA.
ID O81085;
AC O81085;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Cysteine proteinase, An11 precursor.
GN AN11.
OS Ananas comosus (Pineapple).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae; Inceratae sedis;
OC Bromeliaceae; Ananas.
OX NCBI_TaxID=4615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. SMOOTH CAYENNE; TISSUE=STEM;
RA Robertson C.E.; Goodenough P.W.;
RT "Cloning and expression of an ananin gene from pineapple.";
RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AU009830; CA008861.1; -.
DR HSSP; P14080; IYAL.
DR MEROPS; C01.026; -.
DR InterPro; IPR000668; Peptidase_C1.

```

```

DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPAIN.
DR ProDom; PD000158; Peptidase_C1; 1.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW Signal.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 357 CYSTEINE PROTEINASE PRECURSOR, AN11.
SQ SEQUENCE 357 AA; 39606 MW; 0B3930D03E47C31 CRC64;

Query Match
Best Local Similarity 79.0%; Score 83; DB 10; Length 357;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPQSIDWRDSDGAVTSVK 18
DB 123 VPQSIDWRNYGAVTSVK 140

```

Search completed: November 21, 2002, 16:10:00
Job time : 26 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2002, 16:06:44 ; Search time 25 Seconds
(without alignments)
164.838 Million cell updates/sec

Title: US-09-674-738-1

Perfect score: 108

Sequence: 1 VPOSIDMRDYGAVTSVKNN 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp. archaea:*
2: sp. bacteria:*
3: sp. fungi:*
4: sp. human:*
5: sp. invertebrate:*
6: sp. mammal:*
7: sp. mnc:*
8: sp. organelle:*
9: sp. phage:*
10: sp. plant:*
11: sp. rodent:*
12: sp. virus:*
13: sp. vertebrate:*
14: sp. unclassified:*
15: sp. virus:*
16: sp. bacteriophage:*
17: sp. archaea:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	100.0	20	10	Q9S965
2	108	100.0	356	10	Q81084
3	108	100.0	356	10	Q23799
4	99	91.7	351	10	Q23801
5	99	91.7	351	10	Q23791
6	97	89.8	20	10	Q9S8M1
7	93	86.1	20	10	Q9S7J0
8	90	85.2	357	10	Q81085
9	90	83.3	340	10	Q23800
10	90	83.3	352	10	Q24641
11	80	74.1	326	5	Q9SVA7
12	80	74.1	454	5	Q40922
13	79	73.1	339	5	Q26636
14	77	73.1	364	10	Q9S7L5
15	77	71.3	311	5	Q9GRW6
16	77	71.3	334	5	Q9U554

17	77	71.3	337	5	Q9SV59	Q9SV59 delia radic
18	76	70.4	226	10	Q9SMH9	Q9SMH9 carica papa
19	76	70.4	227	10	Q9SMH8	Q9SMH8 carica papa
20	76	70.4	352	10	Q9SMI1	Q9SMI1 carica papa
21	76	70.4	361	10	Q9SMI0	Q9SMI0 carica papa
22	76	70.4	363	10	Q9XGH8	Q9XGH8 nicotiana t
23	76	70.4	363	10	Q43579	Q43579 nicotiana t
24	76	70.4	366	10	Q9AUC5	Q9AUC5 ipomoea bat
25	76	70.4	366	10	Q9W7D5	Q9W7D5 ipomoea bat
26	76	70.4	368	10	Q9W7D6	Q9W7D6 ipomoea bat
27	75	69.4	189	5	Q96677	Q96677 spirometra
28	75	69.4	208	10	Q948S1	Q948S1 daucus caro
29	75	69.4	306	5	Q9GRW4	Q9GRW4 fasciola he
30	75	69.4	326	5	Q9NGW2	Q9NGW2 fasciola gi
31	75	69.4	326	5	Q24944	Q24944 fasciola he
32	75	69.4	326	5	Q9SNT1	Q9SNT1 spirometra
33	75	69.4	326	5	Q9SNT6	Q9SNT6 spirometra
34	75	69.4	360	10	Q9FTI3	Q9FTI3 oryza sativ
35	75	69.4	360	10	Q945R8	Q945R8 sandersonia
36	75	69.4	365	10	Q43580	Q43580 nicotiana t
37	75	69.4	367	10	Q40261	Q40261 mesembryant
38	75	69.4	374	5	Q18455	Q18455 heterodera
39	75	69.4	381	5	Q9GON6	Q9GON6 leishmania
40	75	69.4	394	5	Q9SWR6	Q9SWR6 leishmania
41	75	69.4	443	5	P90628	P90628 leishmania
42	75	69.4	443	5	Q9SWR7	Q9SWR7 leishmania
43	75	69.4	443	5	Q18698	Q18698 leishmania
44	75	69.4	443	5	P92185	P92185 leishmania
45	75	69.4	498	5	O16454	O16454 caenorhabdi

ALIGNMENTS

RESULT 1

ID	Q9S965	PRELIMINARY;	PRT;	20 AA.
AC	Q9S965			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	Bromelain (Fragment).			
OS	Ananas comosus (Pineapple).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae; Inceratae sedis;			
OC	Bromeliaceae; Ananas.			
OX	NCBI_TaxID=4615;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=95298192; PubMed=7779262;			
RA	Harrach T., Eckert K., Schulze-Forster K., Nuck R., Grunow D.,			
RA	Maurer H.R.;			
RT	"Isolation and partial characterization of basic proteinases from stem			
RT	bromelain."			
RU	J. Protein Chem. 14:41-52(1995).			
DR	InterPro: IPR000668; Peptidase.C1.			
DR	Pfam: PF00112; Peptidase.C1.1			
SQ	SEQUENCE 20 AA, 2277 MW; 8F984A1B912C3B8C CRC64;			

Query Match 100.0%; Score 108, DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 7, 1e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	VPOSIDMRDYGAVTSVKNN 20	
Db	1	VPOSIDMRDYGAVTSVKNN 20	
RESULT 2			
ID	Q81084	PRELIMINARY;	PRT; 356 AA.
AC	Q81084		
DT	01-NOV-1998 (TREMBLrel. 08, Created)		

DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 DE Cysteine proteinase, AN8 precursor.
 GN AN8.
 OS Ananas comosus (Pineapple).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae incertae sedis;
 OC Bromeliaceae; Ananas.
 OX NCBI_TaxID=4615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. SMOOTH CAYENNE; TISSUE=STEM;
 RA Robertson C.E., Goodenough P.W.;
 RT "Cloning and expression of an anatin gene from pineapple.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; A009829; CAA08860.1; -;
 DR HSSP; P07711; 1CUL.
 DR MEROPS; C01.005; -;
 DR InterPro; IPR000668; Peptidase_C1.
 DR InterPro; IPR000169; Shprot_acsite.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPA1N.
 DR ProDom; PD000158; Peptidase_C1; 1.
 DR PROSITE; PS00640; THIOI_PROTEASE_ASN; 1.
 DR PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOI_PROTEASE_HIS; UNKNOWN_1.
 KW Hydrolase; Signal; Thiol protease.
 FT SIGNAL
 FT CHAIN 27 356
 SQ SEQUENCE 356 AA; 39528 MW; 97D705C62EC01DB8 CRC64;
 Query Match 100.0%; Score 108; DB 10; Length 356;
 Best Local Similarity 100.0%; Pred. No. 1.8e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VPOSIDWRDYGAVTSVKQN 20
 DB 124 VPOSIDWRDYGAVTSVKQN 143
 RESULT 3
 ID 023799 PRELIMINARY; PRT; 356 AA.
 AC 023799;
 DT 01-JAN-1998 (TReMBLrel. 05, Created)
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 DE FBSB precursor.
 OS Ananas comosus (Pineapple).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae incertae sedis;
 OC Bromeliaceae; Ananas.
 OX NCBI_TaxID=4615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. N67-10; TISSUE=FRUIT;
 RA Mura E., Okamoto Y., Ota S.;
 RT "Cloning and sequencing of cysteine proteinases in Ananas comosus.";
 RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
 DR EMBL; D38532; BAA22544.1; -;
 DR HSSP; P07711; 1CUL.
 DR MEROPS; C01.005; -;
 DR InterPro; IPR001064; Crystallin.
 DR InterPro; IPR000668; Peptidase_C1.
 DR InterPro; IPR000169; Shprot_acsite.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPA1N.
 DR ProDom; PD000158; Peptidase_C1; 1.
 DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
 DR PROSITE; PS00640; THIOI_PROTEASE_ASN; 1.
 DR PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOI_PROTEASE_HIS; UNKNOWN_1.
 KW Hydrolase; Signal; Thiol protease.

FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 123 338 FBSB.
 SQ SEQUENCE 356 AA; 39560 MW; 16C2DCB47238928 CRC64;
 Query Match 100.0%; Score 108; DB 10; Length 356;
 Best Local Similarity 100.0%; Pred. No. 1.8e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VPOSIDWRDYGAVTSVKQN 20
 DB 124 VPOSIDWRDYGAVTSVKQN 143
 RESULT 4
 ID 023801 PRELIMINARY; PRT; 324 AA.
 AC 023801;
 DT 01-JAN-1998 (TReMBLrel. 05, Created)
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 DE FB1035 (Fragment).
 OS Ananas comosus (Pineapple).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae incertae sedis;
 OC Bromeliaceae; Ananas.
 OX NCBI_TaxID=4615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. N67-10; TISSUE=FRUIT;
 RA Mura E., Okamoto Y., Ota S.;
 RT "Cloning and sequencing of cysteine proteinases in Ananas comosus.";
 RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
 DR EMBL; D38534; BAA22546.1; -;
 DR HSSP; P07711; 1CUL.
 DR MEROPS; C01.028; -;
 DR InterPro; IPR000668; Peptidase_C1.
 DR InterPro; IPR000169; Shprot_acsite.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPA1N.
 DR ProDom; PD000158; Peptidase_C1; 1.
 DR PROSITE; PS00640; THIOI_PROTEASE_ASN; 1.
 DR PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOI_PROTEASE_HIS; UNKNOWN_1.
 KW Hydrolase; Thiol protease.
 FT NON TER 1 1
 FT CHAIN 95 312 FB1035.
 SQ SEQUENCE 324 AA; 36038 MW; 057462C59424339F CRC64;
 Query Match 91.7%; Score 99; DB 10; Length 324;
 Best Local Similarity 90.0%; Pred. No. 5e-08;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 VPOSIDWRDYGAVTSVKQN 20
 DB 96 VPOSIDWRDYGAVNEVKQN 115
 RESULT 5
 ID 023791 PRELIMINARY; PRT; 351 AA.
 AC 023791;
 DT 01-JAN-1998 (TReMBLrel. 05, Created)
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 DE Bromelain precursor.
 OS Ananas comosus (Pineapple).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae incertae sedis;
 OC Bromeliaceae; Ananas.
 OX NCBI_TaxID=4615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. N67-10; TISSUE=FRUIT;

RA Muta E., Aramaki H., Takata Y., Kono A., Okamoto Y., Ota S.;
 RT "Cloning and sequencing of fruit bromelain.";
 RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
 DR EMBL: D14059; BAA21649.1; -.
 DR HSSP: P07711; ICUL.

DR MEROPS: C01.028; -.
 DR InterPro: IPR000668; Peptidase_C1.
 DR InterPro: IPR000169; SHprot_acsite.
 DR Pfam: PF00112; Peptidase_C1; 1.
 DR PRINTS: PR00705; PAPA1N_-.
 DR PRODOM: PD000158; Peptidase_C1; 1.
 DR PROSITE: PS00640; THIOI_PROTEASE_ASN; 1.
 DR PROSITE: PS00139; THIOI_PROTEASE_CYS; 1.
 DR PROSITE: PS00639; THIOI_PROTEASE_HIS; UNKNOWN_1.
 KM Hydroxylase; Signal; Thiol protease.
 FT SIGNAL 1 24
 FT CHAIN 122 339 BROMELAIN.
 SQ SEQUENCE 351 AA; 39055 MW; 33781C5144242B0 CRC64;

Query Match 91.7%; Score 99; DB 10; Length 351;
 Best Local Similarity 90.0%; Pred. No. 5,4e-08;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 VPOSIDMRDYGAVTSVKNO 20
 Db 123 VPOSIDMRDYGAVTSVKNO 142

RESULT 6
 Q9S8M1 PRELIMINARY; PRT; 20 AA.

ID Q9S8M1;
 AC Q9S8M1;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE COMOSAIN (fragment).
 OS Ananas comosus (Pineapple).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae; Invertae sedis;
 OC Bromeliaceae; Ananas.
 OC NCBI_TaxID=4615;
 RN (1)
 RP SEQUENCE.
 RX MEDLINE=94330946; PubMed=8053898;
 RA Napper A.D., Bennett S.P., Borowski M., Holdridge M.B., Leonard M.J.,
 RL Rogers E.E., Duan Y., Laursen R.A., Reinhold B., Shames S.L.;
 RT Biochem. J. 301:727-735(1994).
 DR InterPro: IPR000668; Peptidase_C1.
 DR Pfam: PF00112; Peptidase_C1; 1.
 SQ SEQUENCE 20 AA; 2219 MW; 83A84A16232C3EBC CRC64;

Query Match 89.8%; Score 97; DB 10; Length 20;
 Best Local Similarity 94.7%; Pred. No. 4,6e-09;
 Matches 18; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 VPOSIDMRDYGAVTSVKNO 19
 Db 1 VPOSIDMRDYGAVTSVKNO 19

RESULT 7
 Q9S7J0 PRELIMINARY; PRT; 20 AA.
 ID Q9S7J0;
 AC Q9S7J0;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Bromelain (fragment).
 OS Ananas comosus (Pineapple).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae; Invertae sedis;
 OC Bromeliaceae; Ananas.
 OC NCBI_TaxID=4615;

RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95298192; PubMed=7779262;
 RA Harrach T., Eckert K., Schulze-Forster K., Nuck R., Grunow D.,
 RT Maier H.R.;
 RT "Isolation and partial characterization of basic proteinases from stem
 bromelain.";
 RL J. Protein Chem. 14:41-52(1995).
 RN (2)
 RP SEQUENCE.

RX MEDLINE=94330946; PubMed=8053898;
 RA Napper A.D., Bennett S.P., Borowski M., Holdridge M.B., Leonard M.J.,
 RA Rogers E.E., Duan Y., Laursen R.A., Reinhold B., Shames S.L.;
 RL Biochem. J. 301:727-735(1994).
 DR InterPro: IPR000668; Peptidase_C1.
 DR Pfam: PF00112; Peptidase_C1; 1.
 SQ SEQUENCE 20 AA; 2144 MW; 83A842BC812C3EBC CRC64;

Query Match 86.1%; Score 93; DB 10; Length 20;
 Best Local Similarity 94.7%; Pred. No. 2,1e-08;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 VPOSIDMRDYGAVTSVKNO 19
 Db 1 VPOSIDMRDYGAVTSVKNO 19

RESULT 8
 O81085 PRELIMINARY; PRT; 357 AA.

ID O81085;
 AC O81085;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Cysteine proteinase, AN1 precursor.
 GN AN1.
 OS Ananas comosus (Pineapple).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae; Invertae sedis;
 OC Bromeliaceae; Ananas.
 OC NCBI_TaxID=4615;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. SMOOTH CAYENNE; TISSUE=STEM;
 RA Robertson C.E., Goodenough P.W.;
 RT "Cloning and expression of an ananin gene from pineapple.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A7009830; CAA08861.1; -.
 DR HSSP: P14080; IYAL.

DR MEROPS: C01.026; -.
 DR InterPro: IPR000668; Peptidase_C1.
 DR InterPro: IPR000169; SHprot_acsite.
 DR Pfam: PF00112; Peptidase_C1; 1.
 DR PRINTS: PR00705; PAPA1N_-.

DR PRODOM: PD000158; Peptidase_C1; 1.
 DR PROSITE: PS00640; THIOI_PROTEASE_ASN; 1.
 DR PROSITE: PS00639; THIOI_PROTEASE_HIS; UNKNOWN_1.
 KM Signal.
 FT SIGNAL 1 25
 FT CHAIN 27 357 CYSTEINE PROTEINASE PRECURSOR, AN1.
 SQ SEQUENCE 357 AA; 39606 MW; 0B3930D035E47C1 CRC64;

Query Match 85.2%; Score 92; DB 10; Length 357;
 Best Local Similarity 94.4%; Pred. No. 7,8e-07;
 Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 VPOSIDMRDYGAVTSVKNO 18
 Db 123 VPOSIDMRDYGAVTSVKNO 140

RESULT 9
 O23800

Pfam: PF00112; Peptidase_C1; 1.
DR PRINTS; PK00705; PAPAIn.
DR ProDom; PD000158; Peptidase_C1; 1.
DR PROSITE; PS00640; THIOI_PROTEASE_ASN; 1.
DR PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOI_PROTEASE_HIS; UNKNOWN_1.
DR KW Hydrolyase; signal; Thiol protease.
FT SIGNAL 1 24
FT CHAIN 123 340
SEQUENCE 352 AA; 38823 MW; 5969AAC03EBD367 CRC64;

Query Match 83.3%; Score 90; DB 10; Length 352;
Best Local Similarity 85.0%; Pred. No. 1.6e-06;
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0.

QY 1 VQSIDNRDYGVATSVKNQN 20
DB 124 VQSIDNRDYGVATVEKDQN 143
|||||
ID Q95VA7 PRELIMINARY; PRT; 326 AA.
IC Q95VA7
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Cathepsin L.
GN CAT-LIG.
OS Fasciola gigantica.
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Echinostomata; Fascioloidae; Fasciolidae; Fasciola.
NCBI_TaxId=46835;
[1]
RP SEQUENCE FROM N.A.
RN Sobhon P., Weemon K., Grams R., Grams S.V., Korge G., Hofmann A.;
RT "Molecular cloning of cathepsin L encoding genes from Fasciola gigantica."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RR EMBL; AF419329; AAL23917.1; -
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR000169; SHProt_acsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR ProDom; PD000158; Peptidase_C1; 1.
DR PROSITE; PS00640; THIOI_PROTEASE_ASN; UNKNOWN_1.
DR PROSITE; PS00139; THIOI_PROTEASE_CYS; UNKNOWN_1.
DR PROSITE; PS00639; THIOI_PROTEASE_HIS; UNKNOWN_1.
SEQUENCE 326 AA; 37457 MW; 7D5F4AF748E64861 CRC64;

Query Match 74.1%; Score 80; DB 5; Length 326;
Best Local Similarity 78.9%; Pred. No. 6.6e-05;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0.

QY 1 VPQSIDNRDYGVATSVKNQ 19
DB 108 VPASIDNRQYGIVTEKVNQ 126
|||||
ID Q40922 PRELIMINARY; PRT; 454 AA.
AC Q40922; Q40919;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE PSEUDOTZAIN (Cysteine protease).
GN PM33CVSP.
OS Pseudotsuga menziesii (Douglas-fir).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Coniferopsida; Pinaceae; Pseudotsuga.
NCBI_TaxId=3357;
[1]
RP SEQUENCE FROM N.A.

RC STRAIN=[MIRB] FRANCO;
 RX MEDLINE=96269408; PubMed=8682307;
 RA Trambarger T.J., Misra S.;
 RT "Structure and expression of a developmentally regulated cDNA encoding
 RT a cysteine protease (pseudotrypsin) from Douglas fir."; Gene
 RN 112:221-226(1995).
 RP SEQUENCE FROM N.A.
 RC STRAIN=[MIRB] FRANCO;
 RA Trambarger T.J., Misra S.;
 RL Physiol. Plantarum 0:0-0(0).
 RN [3]
 RP SEQUENCE OF 297-347 FROM N.A.
 RC TISSUE=WHOLE SEEDLING;
 RA Trambarger T.J., Misra S.;
 RL Submitted (May-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U41902; AAC94945.1; -
 DR EMBL; 249765; CAA89835.1; -
 DR HSSP; P07711; ICUL.
 DR MEROPS; C01.029; -
 DR InterPro; IPR000118; Granulin.
 DR InterPro; IPR000668; Peptidase_C1.
 DR InterPro; IPR001211; PhospholipaseA2.
 DR InterPro; IPR000169; Shp1ot_acsite.
 DR Pfam; PF00396; Granulin_1.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPAINE.
 DR PRODOM; PD000158; Peptidase_C1; 1.
 DR SMART; SM00277; GRAN_1.
 DR PROSITE; PS00118; PA2_HIS; UNKNOWN_1.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
 DR Hydrolyase; Thiol protease.
 SQ SEQUENCE 454 AA; 49719 MW; 48C1938FD66A023F CRC64;

Query Match 74.1%; Score 80; DB 10; Length 454;
 Best Local Similarity 73.7%; Pred. No. 97e-05;
 Matches 14; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPOSIDMRDYGAVTSVKQ 19
 Db 132 LPESDIMREKGAVTAVKQ 150

RESULT 13
 Q26636 PRELIMINARY; PRT; 339 AA.
 AC Q26636;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Cathepsin L precursor (EC 3.4.22.15).
 OS Sarcophaga peregrina (Flesh fly) (Boettcherisca peregrina).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Oestroidea; Sarcophagidae; Sarcophaga.
 OX NCBI_TaxId=7386;
 RN [1]
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE. AND CHARACTERIZATION.
 RA MEDLINE=9425309; PubMed=8195162;
 RA Homma K.-I., Kurata S., Natori S.;
 RT "Purification, characterization, and cDNA cloning of procathepsin L
 RT from the culture medium of NIH-Sage-4, an embryonic cell line of
 RT Sarcophaga peregrina (Flesh fly), and its involvement in the
 RT differentiation of imaginal discs."; J. Biol. Chem. 269:15258-15264(1994).
 RL J. Biol. Chem. 269:15258-15264(1994).
 CC - FUNCTION: IMPORTANT FOR THE OVERALL DEGRADATION OF PROTEINS IN
 CC LYSOSOMES. REQUIRED FOR DIFFERENTIATION OF IMAGINAL DISKS.
 CC - CATALYTIC ACTIVITY: SPECIFICITY CLOSE TO THAT OF PAPAINE. AS
 CC COMPARED TO CATHEPSIN B, CATHEPSIN L EXHIBITS HIGHER ACTIVITY
 CC TOWARDS PROTEIN SUBSTRATES, BUT HAS LITTLE ACTIVITY ON Z-ARG-ARG-
 CC NMEC, AND NO PEPTIDYL-DIPEPTIDASE ACTIVITY.

CC - SUBUNIT: DIMER OF A HEAVY AND A LIGHT CHAIN LINKED BY DISULFIDE
 CC BONDS (BY SIMILARITY).
 CC - SUBCELLULAR LOCATION: PROCATHEPSIN L IS SECRETED BY IMAGINAL DISKS
 CC WHEN CULTURED IN THE PRESENCE OF 2-HYDROXYECYDSONE. MATURE
 CC CATHEPSIN L IS LYSOSOMAL.
 CC - DEVELOPMENTAL STAGE: HIGHLY EXPRESSED DURING EMBRYONIC DEVELOPMENT
 CC WITH HIGHER LEVELS IN FIRST INSTAR THAN IN THIRD INSTAR.
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE
 CC PAPAINE FAMILY OF THIOL PROTEASES.
 DR EMBL; D16533; BAA03970.1; -
 DR HSSP; P07711; ICUL.
 DR MEROPS; C01.092; -
 DR InterPro; IPR000668; Peptidase_C1.
 DR InterPro; IPR000169; Shp1ot_acsite.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPAINE.
 DR PRODOM; PD000158; Peptidase_C1; 1.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
 KM Hydrolyase; Thiol protease; Glycoprotein; Lysosome; Zymogen; Signal;
 KM Differentiation; Developmental protein.
 FT SIGNAL 1 17
 FT PROPEP 18 121
 FT CHAIN 122 294
 FT PROPEP 295 298
 FT CHAIN 299 339
 FT ACT_SITE 146 146
 FT ACT_SITE 285 285
 FT ACT_SITE 306 306
 FT DISULFID 143 186
 FT DISULFID 177 219
 FT DISULFID 278 328
 FT CARBOHYD 96 96
 SQ SEQUENCE 339 AA; 37847 MW; 7401F3281A2FDA33 CRC64;

Query Match 73.1%; Score 79; DB 5; Length 339;
 Best Local Similarity 68.4%; Pred. No. 0.0001; 1;
 Matches 13; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPOSIDMRDYGAVTSVKQ 19
 Db 122 VPKSVDMREHGAVTGVKQ 140

RESULT 14
 Q9STL5 PRELIMINARY; PRT; 364 AA.
 AC Q9STL5;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Cysteine endopeptidase precursor-like protein.
 GN T99H1_130.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxId=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Choiane N., Robert C., Broctier P., Wincker P., Catcolico L.,
 RA Attienueve F., Saurin W., Weissenbach J., Mewes H.W., Rudd S.,
 RA Lemcke K., Mayer K.F.X., Queller F., Salanoubat M.;
 RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Arabidopsis sequencing project;
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; A0049639; CAB41163.1; -
 DR HSSP; P07711; ICUL.
 DR MEROPS; C01.010; -
 DR InterPro; IPR000886; ER_target.

```

DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPA1N.
DR ProDom; PD000158; Peptidase_C1; 1.
DR PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW Hydrolase; Thiol protease.
SQ SEQUENCE 364 AA; 40971 MW; 81C725D07DCE2E78 CRC64;

Query Match 73.1%; Score 79; DB 10; Length 364;
Best Local Similarity 73.7%; Pred. No. 0.00011;
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKNQ 19
DB 126 VPSSVDVREKGVAVTEVKQ 144
|||:||||:|||||

RESULT 15
O9GRW6 PRELIMINARY; PRT; 311 AA.
AC O9GRW6;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Procathepsin L3 (Fragment).
OS Fasciola hepatica (liver fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Fasciolidae; Fasciola.
OC Echinostomida; Echinostomata; Fascioloidae; Fasciolidae; Fasciola.
OX NCBI_TaxID=6192;
RN [1]
RP SEQUENCE FROM N.A.
RA Harmsen M.M., Cornelissen J.B.W.J., van Milligen F.J., Buijs H.E.C.M.,
RA Jeurissen S.H., Boersma W.;
RT "The propeptide of a cathepsin L-like proteinase induces protection
RT against Fasciola hepatica infection in rats.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; A0279091; CAC12805.1; -.
DR HSSP; P07858; 1HUC.
DR MEROPS; C01.033; -.
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPA1N.
DR ProDom; PD000158; Peptidase_C1; 1.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
FT NON_TER 1
FT CHAIN 92 311 PROCATHEPSIN L3.
SQ SEQUENCE 311 AA; 35409 MW; BB38A5DCB0510845 CRC64;

Query Match 71.3%; Score 77; DB 5; Length 311;
Best Local Similarity 73.7%; Pred. No. 0.0002;
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKNQ 19
DB 93 VPASIDWREYGVVTEVKDQ 111
|||:||||:|||||

Search completed: November 21, 2002, 16:09:59
Job time : 27 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2002, 16:07:14 ; Search time 13 Seconds
(without alignments)

147.899 Million cell updates/sec

Title: US-09-674-738-2

Perfect score: 105

Sequence: 1 VPOSIDWRDSCGAVTSVKNQG 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	100.0	20	2 S46204	ananain (EC 3.4.22
2	105	100.0	345	2 T07839	ananain (EC 3.4.22
3	94	89.5	20	2 S46205	comosain (EC 3.4.2
4	93	88.6	212	2 S03964	stem bromelain (EC
5	93	88.6	356	2 T10514	probable stem brom
6	78	88.6	356	2 T07840	ananain (EC 3.4.22
7	88	83.8	454	2 JC4848	cysteine proteinas
8	84	80.0	324	2 T10518	fruit bromelain (E
9	84	80.0	351	2 T10503	fruit bromelain (E
10	84	80.0	352	2 T09760	chymopapain (EC 3.
11	83	79.0	334	2 AS8195	cathepsin L (EC 3.
12	83	79.0	357	2 T07851	ananain (EC 3.4.22
13	83	79.0	367	2 T12382	cysteine proteinas
14	83	79.0	498	2 T31871	hypothetical prote
15	82	78.1	43	2 S35580	cysteine proteinas
16	82	78.1	214	2 S46476	cysteine proteinas
17	82	78.1	334	1 KHM5L	cathepsin L (EC 3.
18	82	78.1	339	2 A53810	cathepsin L (EC 3.
19	82	78.1	357	2 S49166	cysteine proteinas
20	82	78.1	360	2 T08122	cysteine endopepti
21	81	77.1	361	2 B84601	cysteine proteinas
22	81	77.1	364	2 T06707	cysteine proteinas
23	81	77.1	368	2 JN0718	cysteine proteinas
24	80	76.2	221	2 A59041	cysteine proteinas
25	80	76.2	331	2 JC5441	cathepsin L-like c
26	80	76.2	331	2 JC5442	cathepsin L-like c
27	80	76.2	338	2 JC5443	cathepsin L-like c
28	80	76.2	356	2 A86341	cysteine proteinas
29	80	76.2	361	2 T06708	cysteine proteinas

30	80	76.2	416	2 G86232	cysteine proteinas
31	80	76.2	471	1 KHRZOB	oryzain (EC 3.4.22
32	79	75.2	221	2 A59040	cysteine proteinas
33	79	75.2	334	1 KHR1L	cathepsin L (EC 3.
34	79	75.2	345	1 PPPA	papain (EC 3.4.22.
35	79	75.2	354	2 S25267	cysteine proteinas
36	79	75.2	363	2 S30149	cysteine proteinas
37	79	75.2	371	2 T13022	drought-inducible
38	79	75.2	371	2 T13022	drought-inducible
39	78	74.3	105	2 F84609	probable cysteine
40	78	74.3	218	2 S67481	cathepsin L-like c
41	78	74.3	319	2 T08844	cysteine proteinas
42	78	74.3	358	2 UC787	carrot seed cystei
43	78	74.3	362	1 S12581	cysteine proteinas
44	78	74.3	362	2 S22502	cysteine proteinas
45	78	74.3	365	2 T12040	cysteine proteinas

ALIGNMENTS

RESULT 1
S46204
ananain (EC 3.4.22.31) - pineapple (fragment)
C:Species: Ananas comosus (pineapple)
C:Date: 16-Feb-1995 #sequence_revision 01-Nov-1996 #text_change 07-May-1999
C:Accession: S46204
R:Napier, A.D.; Bennett, S.P.; Borowski, M.; Holdridge, M.B.; Leonard, M.J.C.; Rogers, E.
Biochem. U. 301, 727-735, 1994
A>Title: Purification and characterization of multiple forms of the pineapple-stem-deriv
A:Reference number: S46204; PMID:94330946; PMID:8053898
A:Accession: S46204
A:Molecule type: protein
A:Residues: 1-20 <NAP>
A:Superfamily: papain
C:Keywords: cysteine proteinase; hydrolase.

Query Match 100.0%; Score 105; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.3e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSCGAVTSVKNQG 20
DB 1 VPOSIDWRDSCGAVTSVKNQG 20

RESULT 2
T07839
ananain (EC 3.4.22.31) precursor - pineapple

C:Species: Ananas comosus (pineapple)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 20-Jun-2000
C:Accession: T07839
R:Robertson, C.E.; Goodenough, P.W.
submitted to the EMBL Data Library, November 1997

A:Description: Cloning and expression of ananain gene from pineapple.
A:Reference number: Z16166
A:Accession: T07839

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-345 <ROB>

A:Cross-references: EMBL:AJ002477; PIDN:CAA05487.1
C:Genetics:
A:Gene: ANI

C:Superfamily: papain
C:Keywords: cysteine proteinase; hydrolase

F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-345/Product: ananain #status predicted <MAT>
F:147,279,300/Active site: Cys, His, Asn #status predicted

Query Match 100.0%; Score 105; DB 2; Length 345;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPQSIDWRDMSGAVTSVKNG 20
 |||||
 Db 123 VPQSIDWRDMSGAVTSVKNG 142

RESULT 3
 S46205
 A:Title: comosain (EC 3.4.22.-) - pineapple (fragment)

C:Species: Ananas comosus (pineapple)
 C:Date: 16-Feb-1995 #sequence_revision 01-Nov-1996 #text_change 07-May-1999
 C:Accession: S46205
 R:Nepper, A.D.; Bennett, S.P.; Borowski, M.; Holdridge, M.B.; Leonard, M.J.C.; Rogers, E.
 Biochem. J. 301, 727-735, 1994
 A:Title: Purification and characterization of multiple forms of the pineapple-stem-deriv
 A:Reference number: S46204; MUID:94330946; PMID:805398
 A:Accession: S46205
 A:Molecule type: protein
 A:Residues: 1-20 <NAP>
 C:Superfamily: papain
 C:Keywords: cysteine proteinase; hydrolase

Query Match 89.5%; Score 94; DB 2; Length 20;
 Best Local Similarity 90.0%; Pred. No. 4.2e-09;
 Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPQSIDWRDMSGAVTSVKNG 20
 |||||
 Db 1 VPQSIDWRNYGAVTSVKNG 20

RESULT 4

S03964

A:Title: stem bromelain (EC 3.4.22.32) - pineapple
 C:Species: Ananas comosus (pineapple)
 C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 01-Nov-1996
 C:Accession: S03964
 R:Ritonja, A.; Rowan, A.D.; Buttle, D.J.; Rawlings, N.D.; Turk, V.; Barrett, A.J.
 FEBS Lett. 247, 419-424, 1989
 A:Title: Stem bromelain: amino acid sequence and implications for weak binding of cystat
 A:Reference number: S03964; MUID:89232167; PMID:2714443
 A:Accession: S03964
 A:Molecule type: protein
 A:Residues: 1-212 <RIT>
 C:Superfamily: papain
 C:Keywords: cysteine proteinase; hydrolase

Query Match 88.6%; Score 93; DB 2; Length 212;
 Best Local Similarity 94.7%; Pred. No. 8.8e-08;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPQSIDWRDMSGAVTSVKNG 19
 |||||
 Db 2 VPQSIDWRDYGAVTSVKNG 20

RESULT 5

T10514

A:Title: probable stem bromelain (EC 3.4.22.32) precursor - pineapple
 N:Alternate names: FBSB precursor
 C:Species: Ananas comosus (pineapple)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
 C:Accession: T10514
 R:Mura, E.; Okamoto, Y.; Ota, S.
 submitted to the EMBL Data Library, October 1994
 A:Description: Cloning and sequencing of cysteine proteinases in Ananas comosus.

A:Reference number: Z17065
 A:Accession: T10514
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-356 <MUT>
 A:Cross-references: EMBL:D38532
 A:Experimental source: cv. N67-10; nearly mature fruit
 C:Function:

A:Description: hydrolyzes proteins with broad specificity
 C:Superfamily: papain
 C:Keywords: cysteine proteinase; hydrolase
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-122/Domain: amino-terminal propeptide #status predicted <PRO>
 F:123-338/Product: stem bromelain #status predicted <MAT>
 F:339-356/Domain: carboxyl-terminal propeptide #status predicted <CTP>
 F:148,280,301/Active site: Cys, His, Asn #status predicted

Query Match 88.6%; Score 93; DB 2; Length 356;
 Best Local Similarity 94.7%; Pred. No. 1.6e-07;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPQSIDWRDMSGAVTSVKNG 19
 |||||
 Db 124 VPQSIDWRDYGAVTSVKNG 142

RESULT 6

T07840

A:Title: ananain (EC 3.4.22.31) AN8 precursor - pineapple
 C:Species: Ananas comosus (pineapple)
 C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 20-Jun-2000
 C:Accession: T07840
 R:Robertson, C.E.; Goodenough, P.W.
 submitted to the EMBL Data Library, July 1998
 A:Description: Cloning and expression of ananain gene from pineapple.

A:Reference number: Z16167
 A:Accession: T07840
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-356 <ROB>
 A:Cross-references: EMBL:AJ009829; PIDN:CAA08860.1
 C:Genetics:

C:Superfamily: papain
 C:Keywords: cysteine proteinase; hydrolase
 F:148,280,301/Active site: Cys, His, Asn #status predicted

Query Match 88.5%; Score 93; DB 2; Length 356;
 Best Local Similarity 94.7%; Pred. No. 1.6e-07;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPQSIDWRDMSGAVTSVKNG 19
 |||||
 Db 124 VPQSIDWRDYGAVTSVKNG 142

RESULT 7

JC4848

A:Title: cysteine proteinase (EC 3.4.22.-) - Douglas fir
 N:Alternate names: pseudotzain
 C:Species: Pseudotsuga menziesii (Douglas fir)
 C:Date: 15-Aug-1996 #sequence_revision 15-Oct-1996 #text_change 04-Feb-2000
 C:Accession: JC4848
 R:Traub-Dargatzis, T.J.; Misra, S.
 Gene 172, 221-226, 1996
 A:Title: Structure and expression of a developmentally regulated cDNA encoding a cystein

A:Reference number: JC4848; MUID:96269408; PMID:8682307
 A:Accession: JC4848
 A:Molecule type: mRNA
 A:Residues: 1-454 <TRA>

A:Cross-references: GB:U41902; NID:gl208548; PIDN:AAC49455.1; PID:gl208549
 A:Note: it is uncertain whether Met-1, Met-15 or Met-41 is the initiator
 C:Comment: This enzyme catalyzes the initial stages of storage protein mobilization durin
 C:Superfamily: papain
 C:Keywords: cysteine proteinase; hydrolase
 F:156,292,312/Active site: Cys, His, Asn #status predicted

Query Match 83.8%; Score 88; DB 2; Length 454;
 Best Local Similarity 75.0%; Pred. No. 1.4e-06;
 Matches 15; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPOSIDMRDSCGAVTSYKNG 20
 Db 132 LPESIDMRKGAVTAVKNG 151

RESULT 8

fruit bromelain (EC 3.4.22.33) FB1035 precursor - pineapple (fragment)
 C/Species: Ananas comosus (pineapple)
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
 C/Accession: T10518
 R/Muta: E.; Okamoto, Y.; Ota, S.
 submitted to the EMBL Data Library, October 1994
 A/Description: Cloning and sequencing of cysteine proteinases in Ananas comosus.
 A/Reference number: Z17065
 A/Accession: T10518
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-324 <MUT>
 A/Cross-references: EMBL:D38534
 A/Experimental source: cv. N67-10; nearly mature fruit
 C/Superfamily: papain
 C/Keywords: cysteine proteinase; hydrolase
 F/95-313/Product: fruit bromelain #status predicted <MAT>
 F/314-324/Domain: carboxyl-terminal propeptide #status predicted <CTP>
 F/120,252,273/Active site: Cys, His, Asn #status predicted

Query Match 80.0%; Score 84; DB 2; Length 324;
 Best Local Similarity 84.2%; Pred. No. 4,4e-06;
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VPOSIDMRDSCGAVTSYKNG 19
 Db 96 VPOSIDMRDYGAVNEVKNG 114

RESULT 9

fruit bromelain (EC 3.4.22.33) FB18 precursor - pineapple
 C/Species: Ananas comosus (pineapple)
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
 C/Accession: T10503
 R/Muta: E.; Aramaki, H.; Takata, Y.; Kono, A.; Okamoto, Y.; Ota, S.
 submitted to the EMBL Data Library, January 1993
 A/Description: Cloning and sequencing of fruit bromelain.
 A/Reference number: Z17060
 A/Accession: T10503
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-351 <MUT>
 A/Cross-references: EMBL:D14059
 A/Experimental source: cv. N67-10; nearly mature fruit
 C/Function:
 A/Description: hydrolyses proteins; cysteine proteinase
 C/Superfamily: papain
 C/Keywords: cysteine proteinase; hydrolase
 F/1-24/Domain: signal sequence #status predicted <SIG>
 F/25-131/Domain: amino-terminal propeptide #status predicted <PRO>
 F/132-339/Product: fruit bromelain #status predicted <MAT>
 F/340-351/Domain: carboxyl-terminal propeptide #status predicted <CTP>
 F/117,279,300/Active site: Cys, His, Asn #status predicted

Query Match 80.0%; Score 84; DB 2; Length 351;
 Best Local Similarity 84.2%; Pred. No. 4,4e-06;
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VPOSIDMRDSCGAVTSYKNG 19
 Db 123 VPOSIDMRDYGAVNEVKNG 141

RESULT 10

T09760

chymopapain (EC 3.4.22.6) precursor [validated] - papaya
 C/Species: Carica papaya (papaya)
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 18-Aug-2000
 C/Accession: T09760; S04222; S08285
 R/Comment: I.F. Okamoto, Y. Ota, S.
 submitted to the EMBL Data Library, May 1996

A/Reference number: Z16845
 A/Status: translated from GB/EMBL/DBJ
 A/Accession: T09760
 A/Molecule type: mRNA
 A/Residues: 1-352 <CON>
 A/Cross-references: EMBL:X97789; NID:G1332460; PID:G1332461
 A/Experimental source: leaf
 R/Juquet, A.; Kleinschmidt, T.; Schnek, A.G.; Looze, Y.; Braunitzer, G.
 Biol. Chem. Hoppe-Seyler 370, 425-434, 1989
 A/Title: The thiol proteinases from the latex of Carica papaya L. III. The primary structure of chymopapain.
 A/Reference number: S04222; MUID:89302685; PMID:2500950
 A/Accession: S04222
 A/Molecule type: protein
 A/Residues: 135-352 <JNC>
 R/Watson, D.C.; Yaguchi, M.; Lynn, K.R.
 Biochem. J. 266, 75-81, 1990
 A/Title: The amino acid sequence of chymopapain from Carica papaya.
 A/Reference number: S08285; MUID:90179730; PMID:2106878
 A/Accession: S08285
 A/Molecule type: protein
 A/Residues: 135-352 <MAT>
 C/Superfamily: papain
 C/Keywords: cysteine proteinase; hydrolase
 F/1-26/Domain: signal sequence #status predicted <SIG>
 F/27-134/Domain: propeptide #status predicted <PRO>
 F/135-352/Product: chymopapain #status experimental <MAT>
 F/156-197,190-229,287-338/Disulfide bonds: #status predicted
 F/159,293,313/Active site: Cys, His, Asn #status predicted

Query Match 80.0%; Score 84; DB 2; Length 352;
 Best Local Similarity 84.2%; Pred. No. 4,4e-06;
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 POSIDMRDSCGAVTSYKNG 20
 Db 136 POSIDMRKGAVTAVKNG 154

RESULT 11

A58195
 cathepsin L (EC 3.4.22.15) precursor - pig
 N/Alternate names: major excreted protein (MEP); procathepsin L
 C/Species: Sus scrofa domestica (domestic pig)
 C/Date: 16-Oct-1996 #sequence_revision 18-Oct-1996 #text_change 22-Jun-1999
 C/Accession: A58195; S59911; S59916
 R/Okamura, N.; Tambe, M.; Uchiyama, Y.; Sugita, Y.; Dacheux, F.; Syntin, P.; Dacheux, J.
 Biochim. Biophys. Acta 1245, 221-226, 1995
 A/Title: Direct evidence for the elevated synthesis and secretion of procathepsin L in transgenic mice.
 A/Reference number: S59911; MUID:96069841; PMID:7492581
 A/Accession: A58195
 A/Molecule type: mRNA
 A/Residues: 1-334 <OKA>
 A/Cross-references: DBJ:J37917; NID:G710655; PIDN:BA07140.1; PID:G1468964
 A/Experimental source: epididymis
 A/Note: the authors translated the codon CCG for residue 203 as Thr and CCC for residue 204 as Thr.
 A/Accession: S59911
 A/Molecule type: protein
 A/Residues: 18-31 <OK2>
 C/Function: heterodimer of disulfide linked chains produced from a single chain precursor
 A/Description: catalyzes hydrolysis of peptide bonds in proteins
 A/Pathway: protein degradation
 A/Note: important role in the lysosomal degradation of proteins
 C/Superfamily: papain
 C/Keywords: cysteine proteinase; glycoprotein; heterodimer; hydrolase; lysosome; protein
 F/1-17/Domain: signal sequence #status predicted <SIG>

F:18-113/Domain: propeptide #status predicted <PRO>
 F:114-334/Product: cathepsin L #status predicted <MAT>
 F:135-178,169-212,270-323/Disulfide bonds: #status predicted
 F:138,277,301/Active site: Cys, His, Asn #status predicted
 F:222,292/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 79.0%; Score 83; DB 2; Length 334;
 Best Local Similarity 70.0%; Pred. No. 6.7e-06;
 Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPQSIDWRDSDGAVTSVKNG 20
 ||:|||||:|||||
 Db 114 VPKSVDRKGVTVTKNG 133

RESULT 12

T07851

anainin (EC 3.4.22.31) precursor AN11 - pineapple
 N:Alternate names: cysteine proteinase precursor AN11
 C:Species: Ananas comosus (pineapple)
 C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 20-Jun-2000
 C:Accession: T07851
 R:Robertson, C.E.; Goodenough, P.W.
 submitted to the EMBL Data Library, July 1998
 A:Description: Cloning and expression of anainin gene from pineapple.
 A:Reference number: 216167
 A:Accession: T07851
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-357 <ROB>
 A:Cross-references: EMBL:AJ009830; PIDN:CAA08861.1
 C:Genetics:
 A:Gene: an11
 C:Superfamily: papain
 C:Keywords: cysteine proteinase; hydrolase
 F:1-26/Domain: signal sequence #status predicted <SIG>
 F:27-357/Product: anainin AN11 #status predicted <MAT>
 F:147,281,302/Active site: Cys, His, Asn #status predicted

Query Match 79.0%; Score 83; DB 2; Length 357;
 Best Local Similarity 88.9%; Pred. No. 7.2e-06;
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPQSIDWRDSDGAVTSVK 18
 ||:|||||:|||||
 Db 123 VPQSIDWRNYGAVTSVK 140

RESULT 13

T12382

cysteine proteinase (EC 3.4.22.-) - common ice plant
 N:Alternate names: cysteine endopeptidase; sulphydryl endopeptidase
 C:Species: Mesembryanthemum crystallinum (common ice plant)
 C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 04-Feb-2000
 C:Accession: T12382
 R:Forsthoefel, N.R.; Cushman, J.C.
 submitted to the EMBL Data Library, June 1995
 A:Description: A novel sulphydryl endopeptidase is induced by environmental stress and p
 A:Reference number: 217502
 A:Accession: T12382
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-367 <FOR>
 A:Cross-references: EMBL:U30322; NID:g944915; PID:g944916
 A:Experimental source: leaves NaCl stressed; clone psep7
 C:Genetics:
 A:Gene: Sep1
 C:Superfamily: papain
 C:Keywords: cysteine proteinase; hydrolase
 F:150,287,308/Active site: Cys, His, Asn #status predicted

Query Match 79.0%; Score 83; DB 2; Length 367;
 Best Local Similarity 80.0%; Pred. No. 7.4e-06;

Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 VPQSIDWRDSDGAVTSVKNG 20
 ||:|||||:|||||
 Db 126 VPRSIDWRKGVAVTPVKNG 145

RESULT 14

T31871

hypothetical protein F41E6.6 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T31871
 R:Sammons, L.; Murtagh, J.
 submitted to the EMBL Data Library, July 1997
 A:Description: The sequence of C. elegans cosmid F41E6.
 A:Reference number: 221095
 A:Accession: T31871
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-498 <SAM>
 A:Cross-references: EMBL:AF016448; PIDN:AAB65956.1; GSPDB:GN00023; CESP:F41E6.6
 A:Experimental source: strain Bristol N2; clone F41E6
 C:Genetics:
 A:Gene: CESP:F41E6.6
 A:Map position: 5
 A:Introns: 34/3; 102/3; 140/2; 180/2; 336/2; 357/2; 417/1

Query Match 79.0%; Score 83; DB 2; Length 498;
 Best Local Similarity 70.0%; Pred. No. 1e-05;
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPQSIDWRDSDGAVTSVKNG 20
 ||:|||||:|||||
 Db 264 LPESPDWRKGVAVTQVKNG 283

RESULT 15

S35580

cysteine proteinase (EC 3.4.22.-) IV - mountain papaya (fragment)
 C:Species: Carica pubescens (mountain papaya)
 C:Date: 09-Jun-1994 #sequence_revision 01-Nov-1996 #text_change 07-May-1999
 C:Accession: S35580
 R:Walrave, V.; Jaziri, M.; van Beeumen, J.; Schneek, A.G.; Kleinschmidt, T.; Looze, Y.
 Biol. Chem. Hoppe-Seyler 374, 501-506, 1993
 A:Title: Isolation and preliminary characterization of the cysteine-proteinases from the
 A:Reference number: S35577; MUID:94030669; PMID:8216902
 A:Accession: S35580
 A:Molecule type: protein

A:Residues: 1-43 <WAL>
 A>Note: the source is designated as Carica candamarcensis
 C:Superfamily: papain
 C:Keywords: cysteine proteinase; hydrolase

Query Match 78.1%; Score 82; DB 2; Length 43;
 Best Local Similarity 78.9%; Pred. No. 9.8e-07;
 Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 PQSIDWRDSDGAVTSVKNG 20
 ||:|||||:|||||
 Db 2 PESIDWRKGVAVTPVKNG 20

Search completed: November 21, 2002, 16:10:31
 Job time : 13 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2002, 16:07:14 ; Search time 13 Seconds

(without alignments)
147.899 Million cell updates/sec

Title: US-09-674-738-1

Perfect score: 108

Sequence: 1 VPOSIDMRDYGAVTSVKNON 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 283224 seqs, 9613422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	100.0	212	2 S03964	stem bromelain (EC 3.4.22.32)
2	108	100.0	356	2 T10514	probable stem bromelain (EC 3.4.22.32)
3	108	100.0	356	2 T07840	ananain (EC 3.4.22.32)
4	99	91.7	324	2 T10518	fruit bromelain (EC 3.4.22.32)
5	99	91.7	351	2 T10503	fruit bromelain (EC 3.4.22.32)
6	97	89.8	20	2 S46205	comosain (EC 3.4.22.32)
7	93	86.1	20	2 S46204	ananain (EC 3.4.22.32)
8	93	86.1	345	2 T07839	ananain (EC 3.4.22.32)
9	92	85.2	357	2 T07851	ananain (EC 3.4.22.32)
10	90	83.3	340	2 T10516	fruit bromelain (EC 3.4.22.32)
11	90	83.3	352	2 T10501	fruit bromelain (EC 3.4.22.32)
12	80	74.1	454	2 J04848	cysteine proteinase (EC 3.4.22.32)
13	79	73.1	339	2 A51810	cathepsin L (EC 3.4.22.32)
14	79	73.1	364	2 T06707	cysteine proteinase (EC 3.4.22.32)
15	78	72.2	368	2 T09718	cysteine proteinase (EC 3.4.22.32)
16	76	70.4	352	2 T09760	cysteine proteinase (EC 3.4.22.32)
17	76	70.4	363	2 S30149	cysteine proteinase (EC 3.4.22.32)
18	75	69.4	326	2 T09259	cysteine proteinase (EC 3.4.22.32)
19	75	69.4	334	2 A58195	cathepsin L (EC 3.4.22.32)
20	75	69.4	358	2 J07787	carrot seed cystel
21	75	69.4	365	2 S30150	cysteine proteinase (EC 3.4.22.32)
22	75	69.4	367	2 T12382	hypothetical prote
23	75	69.4	498	2 T31871	cysteine proteinase (EC 3.4.22.32)
24	74	68.5	43	2 S35580	cysteine proteinase (EC 3.4.22.32)
25	74	68.5	214	2 S46476	cathepsin L (EC 3.4.22.32)
26	74	68.5	334	1 KHM5L	cysteine proteinase (EC 3.4.22.32)
27	74	68.5	357	2 S49156	cysteine proteinase (EC 3.4.22.32)
28	74	68.5	360	2 T08122	cysteine proteinase (EC 3.4.22.32)
29	74	68.5	361	2 B84601	cysteine proteinase (EC 3.4.22.32)

30	73	67.6	317	2 S44151	cathepsin L (EC 3.4.22.32)
31	72	66.7	331	2 UC5441	cathepsin L-like c
32	72	66.7	331	2 UC5442	cathepsin L-like c
33	72	66.7	338	2 UC5443	cathepsin L-like c
34	72	66.7	356	2 A85341	cysteine proteinase (EC 3.4.22.32)
35	72	66.7	361	2 T06708	cysteine proteinase (EC 3.4.22.32)
36	72	66.7	416	2 G86232	cysteine proteinase (EC 3.4.22.32)
37	72	66.7	471	1 KHRZOB	oryzain (EC 3.4.22.32)
38	71	65.7	43	2 S35578	cysteine proteinase (EC 3.4.22.32)
39	71	65.7	221	2 A59041	cysteine proteinase (EC 3.4.22.32)
40	71	65.7	231	2 A59040	cysteine proteinase (EC 3.4.22.32)
41	71	65.7	334	1 KHR5L	cysteine proteinase (EC 3.4.22.32)
42	71	65.7	345	1 PPA	papain (EC 3.4.22.32)
43	71	65.7	354	2 S25267	cysteine proteinase (EC 3.4.22.32)
44	71	65.7	364	2 T13022	cysteine proteinase (EC 3.4.22.32)
45	71	65.7	371	2 T13023	drought-inducible

ALIGNMENTS

RESULT 1
S03964
stem bromelain (EC 3.4.22.32) - pineapple
C:Species: Ananas comosus (pineapple)
C>Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 01-Nov-1996
C:Accession: S03964
R:Ritchie, A.; Rowan, A.D.; Buttle, D.J.; Rawlings, N.D.; Turk, V.; Barrett, A.J.
FEBS Lett. 247, 419-424, 1989
A:Title: Stem bromelain: amino acid sequence and implications for weak binding of cystatins
A:Reference number: S03964; PMID:89232167; PMID:2714443
A:Accession: S03964
A:Molecule type: protein
A:Residues: 1-212 <RIT>
C:Superfamily: papain
C:Keywords: cysteine proteinase; hydrolase

Query Match 100.0%; Score 108; DB 2; Length 212;
Best Local Similarity 100.0%; Pred. No. 7.3e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPOSIDMRDYGAVTSVKNON 20
Db 2 VPOSIDMRDYGAVTSVKNON 21

RESULT 2

T10514
probable stem bromelain (EC 3.4.22.32) precursor - pineapple
N:Alternate names: FBSB precursor
C:Species: Ananas comosus (pineapple)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
C:Accession: T10514
R:Muta, E.; Okamoto, Y.; Oca, S.
submitted to the EMBL data library, October 1994
A:Description: Cloning and sequencing of cysteine proteinases in Ananas comosus.
A:Reference number: Z17065
A:Accession: T10514
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-356 <MT>
A:Cross-references: EMBL:D38532
A:Experimental source: cv. N67-10; nearly mature fruit
C:Function:
A:Description: hydrolyzes proteins with broad specificity
C:Superfamily: papain
C:Keywords: cysteine proteinase; hydrolase
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-122/Domain: amino-terminal propeptide #status predicted <PRO>
F:123-336/Product: stem bromelain #status predicted <MT>
F:339-356/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F:148,280,301/Active site: Cys, His, Asn #status predicted

Query Match 100.0%; Score 108; DB 2; Length 356;
 Best Local Similarity 100.0%; Pred. No. 1.3e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPQSIDWRDYGAVTSVKQN 20
 |||||
 Db 124 VPQSIDWRDYGAVTSVKQN 143

RESULT 3
 T07840
 ananain (EC 3.4.22.31) AN8 precursor - pineapple
 C:Species: Ananas comosus (pineapple)
 C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 20-Jun-2000
 C:Accession: T07840
 R:Robertson, C.E.; Goodenough, P.W.
 submitted to the EMBL Data Library, July 1998
 A:Description: Cloning and expression of ananain gene from pineapple.
 A:Reference number: Z16167
 A:Accession: T07840
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: mRNA
 A:Residues: 1-356 <ROB>
 A:Cross-references: EMBL:AJ009829; PIDN:CAA08860.1
 C:Genetics: as8
 A:Gene: an8
 C:Superfamily: papain
 C:Keywords: cysteine proteinase; hydrolase
 F:148,280,301/Active site: Cys, His, Asn #status predicted

Query Match 100.0%; Score 108; DB 2; Length 356;
 Best Local Similarity 100.0%; Pred. No. 1.3e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPQSIDWRDYGAVTSVKQN 20
 |||||
 Db 124 VPQSIDWRDYGAVTSVKQN 143

RESULT 4
 T10518
 fruit bromelain (EC 3.4.22.33) FB1035 precursor - pineapple (fragment)
 C:Species: Ananas comosus (pineapple)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
 C:Accession: T10518
 R:Muta, E.; Okamoto, Y.; Ota, S.
 submitted to the EMBL Data Library, October 1994
 A:Description: Cloning and sequencing of cysteine proteinases in Ananas comosus.
 A:Reference number: Z17065
 A:Accession: T10518
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: mRNA
 A:Residues: 1-324 <MUT>
 A:Cross-references: EMBL:D38534
 A:Experimental source: cv. N67-10; nearly mature fruit
 C:Superfamily: papain
 C:Keywords: cysteine proteinase; hydrolase
 F:95-313/Product: fruit bromelain #status predicted <MAT>
 F:314-324/Domain: carboxyl-terminal propeptide #status predicted <CTP>
 F:120,252,273/Active site: Cys, His, Asn #status predicted

Query Match 91.7%; Score 99; DB 2; Length 324;
 Best Local Similarity 90.0%; Pred. No. 3.3e-08;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPQSIDWRDYGAVTSVKQN 20
 |||||
 Db 96 VPQSIDWRDYGAVTSVKQN 115

RESULT 5
 T10503
 fruit bromelain (EC 3.4.22.33) FB18 precursor - pineapple

C:Species: Ananas comosus (pineapple)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
 C:Accession: T10503
 R:Muta, E.; Aramaki, H.; Takata, Y.; Kono, A.; Okamoto, Y.; Ota, S.
 submitted to the EMBL Data Library, January 1993
 A:Description: Cloning and sequencing of fruit bromelain.
 A:Reference number: Z17060
 A:Accession: T10503
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: mRNA
 A:Residues: 1-351 <MUT>
 A:Cross-references: EMBL:D14059
 A:Experimental source: cv. N67-10; nearly mature fruit
 C:Function: as8
 A:Description: hydrolyses proteins; cysteine proteinase
 C:Superfamily: papain
 C:Keywords: cysteine proteinase; hydrolase
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-121/Domain: amino-terminal propeptide #status predicted <PRO>
 F:122-339/Product: fruit bromelain #status predicted <MAT>
 F:340-351/Domain: carboxyl-terminal propeptide #status predicted <CTP>
 F:147,279,300/Active site: Cys, His, Asn #status predicted

Query Match 91.7%; Score 99; DB 2; Length 351;
 Best Local Similarity 90.0%; Pred. No. 3.6e-08;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPQSIDWRDYGAVTSVKQN 20
 |||||
 Db 123 VPQSIDWRDYGAVTSVKQN 142

RESULT 6
 S46205
 comosain (EC 3.4.22.-) - pineapple (fragment)
 C:Species: Ananas comosus (pineapple)
 C:Date: 16-Feb-1995 #sequence_revision 01-Nov-1996 #text_change 07-May-1999
 C:Accession: S46205
 R:Napper, A.D.; Bennett, S.P.; Borowski, M.; Holdridge, M.B.; Leonard, M.J.C.; Rogers, E.
 Biochem. J. 301, 727-735, 1994
 A:Title: Purification and characterization of multiple forms of the pineapple-stem-deriv.
 A:Reference number: S46204; MUID:94330946; PMID:8053898
 A:Accession: S46205
 A:Molecule type: protein
 A:Residues: 1-20 <NAP>
 A:Superfamily: papain
 C:Keywords: cysteine proteinase; hydrolase

Query Match 89.8%; Score 97; DB 2; Length 20;
 Best Local Similarity 94.7%; Pred. No. 3.1e-09;
 Matches 18; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPQSIDWRDYGAVTSVKQN 19
 |||||
 Db 1 VPQSIDWRDYGAVTSVKQN 19

RESULT 7
 S46204
 ananain (EC 3.4.22.31) - pineapple (fragment)
 C:Species: Ananas comosus (pineapple)
 C:Date: 16-Feb-1995 #sequence_revision 01-Nov-1996 #text_change 07-May-1999
 C:Accession: S46204
 R:Napper, A.D.; Bennett, S.P.; Borowski, M.; Holdridge, M.B.; Leonard, M.J.C.; Rogers, E.
 Biochem. J. 301, 727-735, 1994
 A:Title: Purification and characterization of multiple forms of the pineapple-stem-deriv.
 A:Reference number: S46204; MUID:94330946; PMID:8053898
 A:Accession: S46204
 A:Molecule type: protein
 A:Residues: 1-20 <NAP>
 C:Superfamily: papain
 C:Keywords: cysteine proteinase; hydrolase

Query Match 86.1%; Score 93; DB 2; Length 20;
 Best Local Similarity 94.7%; Pred. No. 1.4e-08;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPOSIDWRDYGAVTSVKNQ 19
 |||||
 Db 1 VPOSIDWRDYGAVTSVKNQ 19

RESULT 8

ananaIn (EC 3.4.22.31) precursor - pineapple
 C/Species: Ananas comosus (pineapple)
 C/Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 20-Jun-2000
 C/Accession: T07839
 R/Robertson, C.E.; Goodenough, P.W.
 submitted to the EMBL Data Library, November 1997
 A/Description: Cloning and expression of ananain gene from pineapple.
 A/Reference number: Z16166
 A/Accession: T07839
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-345 <ROB>
 A/Cross-references: EMBL:AJ002477; PIDN:CAA05487.1
 C/Genetics:
 A/Gene: ANI
 A/Superfamily: papain
 C/Keywords: cysteine proteinase; hydrolase
 F/1-24/Domain: signal sequence #status predicted <SIG>
 F/25-345/Product: ananain #status predicted <MAT>
 F/147,279,300/Active site: Cys, His, Asn #status predicted

Query Match 86.1%; Score 93; DB 2; Length 345;
 Best Local Similarity 94.7%; Pred. No. 3.3e-07;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPOSIDWRDYGAVTSVKNQ 19
 |||||
 Db 123 VPOSIDWRDYGAVTSVKNQ 141

RESULT 9

ananaIn (EC 3.4.22.31) precursor ANI1 - pineapple
 N/Alternate names: cysteine proteinase precursor ANI1
 C/Species: Ananas comosus (pineapple)
 C/Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 20-Jun-2000
 C/Accession: T07851
 R/Robertson, C.E.; Goodenough, P.W.
 submitted to the EMBL Data Library, July 1998
 A/Description: Cloning and expression of ananain gene from pineapple.
 A/Reference number: Z16167
 A/Accession: T07851
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-357 <ROB>
 A/Cross-references: EMBL:AJ009830; PIDN:CAA08861.1
 C/Genetics:
 A/Gene: anil
 A/Superfamily: papain
 C/Keywords: cysteine proteinase; hydrolase
 F/1-26/Domain: signal sequence #status predicted <SIG>
 F/27-357/Product: ananain ANI1 #status predicted <MAT>
 F/147,281,302/Active site: Cys, His, Asn #status predicted

Query Match 85.2%; Score 92; DB 2; Length 357;
 Best Local Similarity 94.4%; Pred. No. 5e-07;
 Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPOSIDWRDYGAVTSVKN 18
 |||||
 Db 123 VPOSIDWRDYGAVTSVKN 140

RESULT 10
 T10516
 fruit bromelain (EC 3.4.22.33) FB22 precursor - pineapple (fragment)

C/Species: Ananas comosus (pineapple)
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
 C/Accession: T10516
 R/Muta, E.; Okamoto, Y.; Ota, S.
 submitted to the EMBL Data Library, October 1994
 A/Description: Cloning and sequencing of cysteine proteinases in Ananas comosus.
 A/Reference number: Z17065
 A/Accession: T10516
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-340 <MUT>
 A/Cross-references: EMBL:D38533
 A/Experimental source: cv. N67-10; nearly mature fruit
 C/Function:
 A/Description: hydrolyzes proteins with broad specificity; cysteine proteinase
 C/Superfamily: papain
 C/Keywords: cysteine proteinase; hydrolase
 F/1-24/Domain: signal sequence #status predicted <SIG>
 F/25-121/Domain: amino-terminal propeptide #status predicted <PRO>
 F/122-339/Product: fruit bromelain #status predicted <MAT>
 F/147,280,301/Active site: Cys, His, Asn #status predicted

Query Match 83.3%; Score 90; DB 2; Length 340;
 Best Local Similarity 85.0%; Pred. No. 9.9e-07;
 Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VPOSIDWRDYGAVTSVKNQ 20
 |||||
 Db 123 VPOSIDWRDYGAVTEVDON 142

RESULT 11

T10501
 fruit bromelain (EC 3.4.22.33) FB13 precursor - pineapple
 C/Species: Ananas comosus (pineapple)
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
 C/Accession: T10501
 R/Muta, E.; Aramaki, H.; Takata, Y.; Kono, A.; Okamoto, Y.; Ota, S.
 submitted to the EMBL Data Library, January 1993
 A/Description: Cloning and sequencing of fruit bromelain.
 A/Reference number: Z17060
 A/Accession: T10501
 A/Status: translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-352 <MUT>
 A/Cross-references: EMBL:D14058
 A/Experimental source: cv. N67-10; fruit; nearly mature
 C/Function:
 A/Description: hydrolyzes proteins; cysteine proteinase
 C/Superfamily: papain
 C/Keywords: cysteine proteinase; hydrolase
 F/1-24/Domain: signal sequence #status predicted <SIG>
 F/25-122/Domain: amino-terminal propeptide #status predicted <PRO>
 F/123-340/Product: fruit bromelain #status predicted <MAT>
 F/341-352/Domain: carboxyl-terminal propeptide #status predicted <CTP>
 F/148,281,302/Active site: Cys, His, Asn #status predicted

Query Match 83.3%; Score 90; DB 2; Length 352;
 Best Local Similarity 85.0%; Pred. No. 1e-06;
 Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VPOSIDWRDYGAVTSVKNQ 20
 |||||
 Db 124 VPOSIDWRDYGAVTEVDON 143

RESULT 12
 J048948
 cysteine proteinase (EC 3.4.22.-) - Douglas fir

Search completed: November 21, 2002, 16:10:31
Job time : 14 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2002, 16:07:39 ; Search time 11.5 Seconds
(Without alignments)
51.170 Million cell updates/sec

Title: US-09-674-738-1

Perfect score: 108

Sequence: 1 VPOSIDMRDYGAVTSVKNON 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 2942292 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
1: /cgcn2_6/ptodata/1/aa/5A.COMB.pep.*
2: /cgcn2_6/ptodata/1/aa/5B.COMB.pep.*
3: /cgcn2_6/ptodata/1/aa/6A.COMB.pep.*
4: /cgcn2_6/ptodata/1/aa/6B.COMB.pep.*
5: /cgcn2_6/ptodata/1/aa/6C.COMB.pep.*
6: /cgcn2_6/ptodata/1/aa/6D.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	100.0	213	2	US-08-578-701A-1
2	108	100.0	213	2	US-08-360-693-1
3	88	81.5	18	4	US-09-382-689A-1
4	79	73.1	127	4	US-09-325-932A-176
5	79	73.1	148	4	US-09-325-932A-167
6	79	73.1	278	4	US-09-325-932A-165
7	79	73.1	339	3	US-09-120-365-62
8	79	73.1	339	3	US-09-120-365-80
9	79	73.1	339	4	US-09-515-039-62
10	79	73.1	339	4	US-09-515-039-60
11	75	69.4	334	3	US-08-883-526-4
12	75	69.4	443	3	US-09-120-365-59
13	75	69.4	443	4	US-09-515-039-69
14	74	68.5	334	3	US-09-120-365-64
15	74	68.5	334	4	US-09-515-039-64
16	73	67.6	336	4	US-09-325-932A-153
17	72	66.7	129	4	US-09-325-932A-159
18	72	66.7	191	3	US-08-906-769-178
19	72	66.7	191	3	US-08-906-616-178
20	72	66.7	191	4	US-08-639-075A-178
21	72	66.7	191	4	US-09-004-731-77
22	72	66.7	191	4	US-09-012-431-178
23	72	66.7	191	4	US-08-749-699-77
24	72	66.7	191	4	US-09-013-692-178
25	72	66.7	191	4	US-08-906-613-178
26	72	66.7	191	4	US-09-004-729-77
27	72	66.7	220	4	US-09-004-731-95

28	72	66.7	220	4	US-08-749-699-95	Sequence 95, Appl
29	72	66.7	220	4	US-09-004-729-95	Sequence 95, Appl
30	72	66.7	226	4	US-09-004-731-8	Sequence 8, Appl
31	72	66.7	226	4	US-08-749-699-8	Sequence 8, Appl
32	72	66.7	226	4	US-09-004-729-8	Sequence 8, Appl
33	72	66.7	248	4	US-09-325-932A-161	Sequence 161, Appl
34	72	66.7	254	4	US-09-004-731-89	Sequence 89, Appl
35	72	66.7	254	4	US-09-004-731-92	Sequence 92, Appl
36	72	66.7	254	4	US-08-749-699-89	Sequence 89, Appl
37	72	66.7	254	4	US-08-749-699-92	Sequence 92, Appl
38	72	66.7	254	4	US-09-004-729-89	Sequence 89, Appl
39	72	66.7	254	4	US-09-004-729-92	Sequence 92, Appl
40	72	66.7	327	4	US-09-004-731-2	Sequence 2, Appl
41	72	66.7	327	4	US-09-004-731-5	Sequence 5, Appl
42	72	66.7	327	4	US-08-749-699-2	Sequence 2, Appl
43	72	66.7	327	4	US-08-749-699-5	Sequence 5, Appl
44	72	66.7	327	4	US-09-004-729-2	Sequence 2, Appl
45	72	66.7	327	4	US-09-004-729-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-08-578-701A-1
; Sequence 1, Application US/08578701A
; Patent No. 5824305
; GENERAL INFORMATION:
; APPLICANT: Myro, Tracy L.
; TITLE OF INVENTION: MEDICAL USE OF STEW BROMELAIN PROTEASE
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/578,701A
; FILING DATE: 22-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Baldwin, Geraldine F.
; REGISTRATION NUMBER: 31,232
; REFERENCE/DOCKET NUMBER: 7116-046
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO. 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 213 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-578-701A-1

Query Match 100.0%; Score 108; DB 2; Length 213;
Best Local Similarity 100.0%; Pred. No. 9.5e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY 1 VPOSIDMRDYGAVTSVKNON 20
DB 2 VPOSIDMRDYGAVTSVKNON 21

```
RESULT 2
US-08-360-693-1
; Sequence 1, Application US/08360693
; Patent No. 5928640
; GENERAL INFORMATION:
; APPLICANT: MYNOTT, Tracey L.
; TITLE OF INVENTION: USE OF ENZYMES, ESPECIALLY BROMELAIN, IN
; THE TREATMENT OF DIARRHOEA
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/360,693
; FILING DATE: 14-FEB-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/GB93/01374
; FILING DATE: 30-JUN-1993
; PRIOR APPLICATION NUMBER: GB 931389.4
; FILING DATE: 25-JUN-1993
; APPLICATION DATA:
; APPLICATION NUMBER: GB 9308164.4
; FILING DATE: 20-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9213862.7
; FILING DATE: 30-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30357/115/KIST
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 213 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-360-693-1
Query Match 100.0%; Score 108; DB 2; Length 213;
Best Local Similarity 100.0%; Pred. No. 9.5e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKNN 20
Db 2 VPOSIDWRDYGAVTSVKNN 21

RESULT 3
US-09-382-689A-1
; Sequence 1, Application US/09382689A
; Patent No. 6335427
; GENERAL INFORMATION:
; APPLICANT: MYNOTT, Tracy Lehanne
; APPLICANT: Engwerda, Christian
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.076001
; CURRENT APPLICATION NUMBER: US/09/382,689A

Query Match 73.1%; Score 79; DB 4; Length 127;
Best Local Similarity 68.4%; Pred. No. 4.3e-06;
Matches 13; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKNN 19
Db 67 LPKDFDWRDHGAVTAVKNN 85

RESULT 4
US-09-325-932A-176
; Sequence 176, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; TITLE OF INVENTION: Compositions affecting programmed cell
; death and their use in the modification of forestry plant develop-
; ment
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 176
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
; US-09-325-932A-176
Query Match 81.5%; Score 88; DB 4; Length 18;
Best Local Similarity 88.9%; Pred. No. 1.1e-08;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKNN 18
Db 1 VPOSIDWRDYGAVTSVKNN 18

RESULT 5
US-09-325-932A-167
; Sequence 167, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; TITLE OF INVENTION: Compositions affecting programmed cell
; death and their use in the modification of forestry plant develop-
; ment
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 167
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
; US-09-325-932A-167
```


US-09-325-932A-167

Query Match 73.1% Score 79; DB 4; Length 148;
Best Local Similarity 68.4%; Pred. No. 5.2e-06;
Matches 13; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPQSDWRDYGAVTSVKNO 19
||:||||:||||:|
DB 57 LPKDFWRDHGAVTGVKQ 75

RESULT 6
US-09-325-932A-165
; Sequence 165, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Fliin, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of forestry plant devel
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: PaetsEQ for Windows Version 3.0
; SEQ ID NO 165
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
US-09-325-932A-165

Query Match 73.1% Score 79; DB 4; Length 278;
Best Local Similarity 68.4%; Pred. No. 1.2e-05;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPQSDWRDYGAVTSVKNO 19
||:||||:||||:|
DB 144 LPDFWRDHGAVTGVKQ 162

RESULT 7
US-09-120-365-62
; Sequence 62, Application US/09120365
; Patent No. 6103514
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/120,365
; CURRENT FILING DATE: 1998-07-22
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Sarcophaga cathepsin L
US-09-120-365-62

Query Match 73.1% Score 79; DB 3; Length 339;
Best Local Similarity 68.4%; Pred. No. 1.5e-05;
Matches 13; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPQSDWRDYGAVTSVKNO 19
||:||||:||||:|
DB 122 VPKSDWRDHGAVTGVKQ 140

RESULT 8
US-09-120-365-80
; Sequence 80, Application US/09120365
; Patent No. 6103514

; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/120,365
; CURRENT FILING DATE: 1998-07-22
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Sarcophaga cathepsin L
US-09-120-365-80

Query Match 73.1% Score 79; DB 3; Length 339;
Best Local Similarity 68.4%; Pred. No. 1.5e-05;
Matches 13; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPQSDWRDYGAVTSVKNO 19
||:||||:||||:|
DB 122 VPKSDWRDHGAVTGVKQ 140

RESULT 9
US-09-515-039-62
; Sequence 62, Application US/09515039
; Patent No. 6214599
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/515,039
; CURRENT FILING DATE: 2000-03-06
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Sarcophaga cathepsin L
US-09-515-039-62

Query Match 73.1% Score 79; DB 4; Length 339;
Best Local Similarity 68.4%; Pred. No. 1.5e-05;
Matches 13; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPQSDWRDYGAVTSVKNO 19
||:||||:||||:|
DB 122 VPKSDWRDHGAVTGVKQ 140

RESULT 10
US-09-515-039-80
; Sequence 80, Application US/09515039
; Patent No. 6214599
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/515,039
; CURRENT FILING DATE: 2000-03-06
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Sarcophaga cathepsin L
US-09-515-039-80

Query Match 73.1%; Score 79; DB 4; Length 339;
 Best Local Similarity 68.4%; Pred. No. 1.5e-05;
 Matches 13; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPQSIDWRDYGAVTSVKNQ 19
 Db 122 VPKSVDMREHGAVTGVKQ 140

RESULT 11
 US-08-883-526-4
 ; Sequence 4, Application US/08883526
 ; Patent No. 6033893
 ; GENERAL INFORMATION:
 ; APPLICANT: Bandman, Olga
 ; APPLICANT: Guegler, Karl J.
 ; APPLICANT: Corley, Neil C.
 ; APPLICANT: Shah, Purvi
 ; TITLE OF INVENTION: NEW HUMAN CATHEPSIN
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/883,526
 ; FILING DATE: Herewith
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0331 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-855-0555
 ; TELEFAX: 415-845-4166
 ; TELEX:

INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 334 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: 1468964
 ; CLONE: Genbank
 ; US-08-883-526-4

Query Match 69.4%; Score 75; DB 3; Length 334;
 Best Local Similarity 68.4%; Pred. No. 7e-05;
 Matches 13; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPQSIDWRDYGAVTSVKNQ 19
 Db 114 VPKSVDMREHGAVTGVKQ 132

RESULT 12
 US-09-120-365-69
 ; Sequence 69, Application US/09120365
 ; Patent No. 6103514
 ; GENERAL INFORMATION:

; APPLICANT: Natori, Shunji
 ; TITLE OF INVENTION: NEW PROTEASE
 ; FILE REFERENCE: 32290-144749
 ; CURRENT APPLICATION NUMBER: US/09/120,365
 ; CURRENT FILING DATE: 1998-07-22
 ; EARLIER APPLICATION NUMBER: JP 9-333 474
 ; EARLIER FILING DATE: 1997-11-18
 ; NUMBER OF SEQ ID NOS: 101
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 69
 ; LENGTH: 443
 ; TYPE: PRT
 ; ORGANISM: Leishmania
 ; US-09-120-365-69

Query Match 69.4%; Score 75; DB 3; Length 443;
 Best Local Similarity 68.4%; Pred. No. 9.9e-05;
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPQSIDWRDYGAVTSVKNQ 19
 Db 126 VPDVDMREKGAVTGVKQ 144

RESULT 13
 US-09-515-039-69
 ; Sequence 69, Application US/09515039
 ; Patent No. 6214599
 ; GENERAL INFORMATION:
 ; APPLICANT: Natori, Shunji
 ; TITLE OF INVENTION: NEW PROTEASE
 ; FILE REFERENCE: 32290-144749
 ; CURRENT APPLICATION NUMBER: US/09/515,039
 ; CURRENT FILING DATE: 2000-03-06
 ; EARLIER APPLICATION NUMBER: JP 9-333 474
 ; EARLIER FILING DATE: 1997-11-18
 ; NUMBER OF SEQ ID NOS: 101
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 69
 ; LENGTH: 443
 ; TYPE: PRT
 ; ORGANISM: Leishmania
 ; US-09-515-039-69

Query Match 69.4%; Score 75; DB 4; Length 443;
 Best Local Similarity 68.4%; Pred. No. 9.9e-05;
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPQSIDWRDYGAVTSVKNQ 19
 Db 126 VPDVDMREKGAVTGVKQ 144

RESULT 14
 US-09-120-365-64
 ; Sequence 64, Application US/09120365
 ; Patent No. 6103514
 ; GENERAL INFORMATION:
 ; APPLICANT: Natori, Shunji
 ; TITLE OF INVENTION: NEW PROTEASE
 ; FILE REFERENCE: 32290-144749
 ; CURRENT APPLICATION NUMBER: US/09/120,365
 ; CURRENT FILING DATE: 1998-07-22
 ; EARLIER APPLICATION NUMBER: JP 9-333 474
 ; EARLIER FILING DATE: 1997-11-18
 ; NUMBER OF SEQ ID NOS: 101
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 64
 ; LENGTH: 334
 ; TYPE: PRT
 ; ORGANISM: Murine cathepsin L
 ; US-09-120-365-64

Query Match	68.5%	Score 74	DB 3	Length 334
Best Local Similarity	63.2%	Pred. No. 0.0001		
Matches 12; Conservative	4	Mismatches 3	Indels 0	Gaps 0

```

OY      1  VPOSIDWRDYGAVTSVKNO 19
          :|:|:|:|:|:|:|:|:|:|
Db      114 1PKSVDWREKGCVTVPVKNO 132

```

```

RESULT 15
US-09-515-039-64
: Sequence 64, Application US/09515039
: Patent No. 6214599
: GENERAL INFORMATION:
: APPLICANT: Natori, Shunji
: TITLE OF INVENTION: NEW PROTEASE
: FILE REFERENCE: 32290-144749
: CURRENT APPLICATION NUMBER: US/09/515, 039
: CURRENT FILING DATE: 2000-03-06
: EARLIER APPLICATION NUMBER: JP 9-333 474
: EARLIER FILING DATE: 1997-11-18
: NUMBER OF SEQ ID NOS: 101
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 64
: LENGTH: 334
: TYPE: PRT
: ORGANISM: Murine cathepsin L
US-09-515-039-64

```

Query Match	68.5%	Score 74;	DB 4;	Length 334;
Best Local Similarity	63.2%;	Pred. No. 0.0001;		
Matches 12; Conservative	4;	Mismatches 3;	Indels 0;	Gaps 0;

```
QY      1 VPQSIDWRDYGAVTSVKNQ 19
      :|:|:|:|:| | | | |
Db     114 IPKSVDWREKGCVTPIKNQ 132
```

Search completed: November 21, 2002, 16:11:01
Job time : 12.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2002, 16:07:39 ; Search time 11.5 Seconds
(without alignments)
51.170 Million cell updates/sec

Title: US-09-674-738-2

Perfect score: 105
Sequence: 1 VPOSIDWRDGAVTSVKNQG 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents RA:

1: /cgn2_6/ptodata/1/aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/aa/6CTUS.COMB.pep.*
6: /cgn2_6/ptodata/1/aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	93	88.6	213	2	US-08-578-701A-1
2	93	88.6	213	2	US-08-360-693-1
3	83	79.0	334	3	US-08-883-526-4
4	83	79.0	443	3	US-09-120-365-69
5	83	79.0	443	4	US-09-515-039-69
6	82	78.1	127	4	US-09-325-932A-176
7	82	78.1	148	4	US-09-325-932A-167
8	82	78.1	278	4	US-09-325-932A-165
9	82	78.1	334	3	US-09-120-365-64
10	82	78.1	334	4	US-09-515-039-64
11	82	78.1	339	3	US-09-120-365-62
12	82	78.1	339	3	US-09-120-365-80
13	82	78.1	339	4	US-09-515-039-62
14	82	78.1	339	4	US-09-515-039-80
15	81	77.1	336	4	US-09-325-932A-153
16	80	76.2	129	4	US-09-325-932A-159
17	80	76.2	442	4	US-08-821-994-66
18	80	76.2	443	4	US-08-821-994-65
19	79	75.2	18	4	US-09-382-689A-1
20	79	75.2	212	4	US-08-860-255A-4
21	79	75.2	225	4	US-08-821-994-51
22	79	75.2	345	3	US-09-120-365-73
23	79	75.2	345	4	US-09-515-039-73
24	79	75.2	345	4	US-08-827-171B-7
25	79	75.2	370	4	US-09-325-932A-169
26	79	75.2	374	4	US-08-821-994-67
27	79	75.2	374	4	US-08-821-994-69

28	78	74.3	216	1	US-08-546-712-2	Sequence 2, Appl
29	78	74.3	216	2	US-08-751-105-2	Sequence 2, Appl
30	78	74.3	359	2	US-08-560-398-10	Sequence 10, Appl
31	78	74.3	362	3	US-09-120-365-74	Sequence 74, Appl
32	78	74.3	362	4	US-09-515-039-74	Sequence 74, Appl
33	77	73.3	302	4	US-09-325-932A-148	Sequence 148, Appl
34	77	73.3	326	4	US-09-120-365-67	Sequence 67, Appl
35	77	73.3	326	4	US-09-515-039-67	Sequence 67, Appl
36	77	73.3	333	1	US-08-208-007A-9	Sequence 9, Appl
37	77	73.3	333	1	US-08-330-121B-5	Sequence 5, Appl
38	77	73.3	333	3	US-08-883-526-3	Sequence 3, Appl
39	77	73.3	333	3	US-09-120-365-65	Sequence 65, Appl
40	77	73.3	333	4	US-09-515-039-65	Sequence 65, Appl
41	77	73.3	333	4	US-08-860-255A-3	Sequence 3, Appl
42	77	73.3	333	4	US-08-915-095A-9	Sequence 9, Appl
43	77	73.3	333	4	US-08-798-096A-9	Sequence 9, Appl
44	77	73.3	333	4	US-08-798-095A-9	Sequence 9, Appl
45	77	73.3	333	5	PCT-US95-13820-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-08-578-701A-1
Sequence 1, Application US/08578701A
Patent No. 5824305
GENERAL INFORMATION:
APPLICANT: Myroth, Tracy L.
TITLE OF INVENTION: MEDICAL USE OF STEM BROMELAIN PROTEASE
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESSES:
ADDRESS: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/578,701A
FILING DATE: 22-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7116-046
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 213 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-578-701A-1

Query Match 88.6%; Score 93; DB 2; Length 213;
Best Local Similarity 94.7%; Pred. No. 1.1e-07;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 VPOSIDWRDGAVTSVKNQ 19
|||
Db 2 VPOSIDWRDGAVTSVKNQ 20

RESULT 2

US-08-360-693-1
 ; Sequence 1, Application US/08360693
 ; Patent No. 5928640
 ; GENERAL INFORMATION:
 ; APPLICANT: MYNOTT, Tracey L.
 ; TITLE OF INVENTION: USE OF ENZYMES, ESPECIALLY BROMELAIN, IN
 ; THE TREATMENT OF DIARRHOEA
 ; NUMBER OF SEQUENCES: 1
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/360,693
 ; FILING DATE: 14-FEB-1995
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: WO PCT/GB93/01374
 ; FILING DATE: 30-JUN-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 931389.4
 ; FILING DATE: 25-JUN-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9308164.4
 ; FILING DATE: 20-APR-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9213862.7
 ; FILING DATE: 30-JUN-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BENT, Stephen A.
 ; REGISTRATION NUMBER: 29,768
 ; REFERENCE/DOCKET NUMBER: 30357/115/KIST
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)672-5300
 ; TELEFAX: (202)672-5399
 ; TELEX: 904136
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 213 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-360-693-1

Query Match 88.6%; Score 93; DB 2; Length 213;
 Best Local Similarity 94.7%; Pred. No. 1.1e-07;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VPOSIDWRDYGAVTSVKNO 19
 |||||
 Db 2 VPOSIDWRDYGAVTSVKNO 20

RESULT 3

US-08-883-526-4
 ; Sequence 4, Application US/08883526
 ; Patent No. 6033893
 ; GENERAL INFORMATION:
 ; APPLICANT: Bandman, Olga
 ; APPLICANT: Guegler, Karl J.
 ; APPLICANT: Corley, Neil C.
 ; APPLICANT: Shah, Purvi
 ; TITLE OF INVENTION: NEW HUMAN CATHEPSIN
 ; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/883,526
 ; FILING DATE: Herewith
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0331 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-855-0555
 ; TELEFAX: 415-845-4166
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 334 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: 1468964
 ; CLONE: GenBank
 ; US-08-883-526-4

Query Match 79.0%; Score 83; DB 3; Length 334;
 Best Local Similarity 70.0%; Pred. No. 7.8e-06;
 Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VPOSIDWRDYGAVTSVKNO 20
 |||||
 Db 114 VPKSVDMREKGYVTAVKNO 133

RESULT 4

US-09-120-365-69
 ; Sequence 69, Application US/09120365
 ; Patent No. 6103514
 ; GENERAL INFORMATION:
 ; APPLICANT: Natori, Shunji
 ; TITLE OF INVENTION: NEW PROTEASE
 ; FILE REFERENCE: 32290-144749
 ; CURRENT APPLICATION NUMBER: US/09/120,365
 ; CURRENT FILING DATE: 1998-07-22
 ; EARLIER APPLICATION NUMBER: JP 9-333 474
 ; EARLIER FILING DATE: 1997-11-18
 ; NUMBER OF SEQ ID NOS: 101
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 69
 ; LENGTH: 443
 ; TYPE: PRT
 ; ORGANISM: Leishmania
 ; US-09-120-365-69

Query Match 79.0%; Score 83; DB 3; Length 443;
 Best Local Similarity 70.0%; Pred. No. 1.1e-05;
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VPOSIDWRDYGAVTSVKNO 20
 |||||

Db 126 VPDVDMREKAVTPVKNKG 145

RESULT 5
US-09-515-039-69
; Sequence 69, Application US/09515039
; Patent No. 6214599
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/515.039
; CURRENT FILING DATE: 2000-03-06
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 69
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Leishmania
US-09-515-039-69

Query Match 79.0%; Score 83; DB 4; Length 443;
Best Local Similarity 70.0%; Pred. No. 1.1e-05;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VPOSIDMRDGAATSVKNKG 20
Db 126 VPDVDMREKAVTPVKNKG 145

RESULT 6
US-09-325-932A-176
; Sequence 176, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325.932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 176
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
US-09-325-932A-176

Query Match 78.1%; Score 82; DB 4; Length 127;
Best Local Similarity 70.0%; Pred. No. 3.6e-06;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VPOSIDMRDGAATSVKNKG 20
Db 67 LPKDFMRDHGAVTAVVKNKG 86

RESULT 7
US-09-325-932A-167
; Sequence 167, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325.932A
; CURRENT FILING DATE: 1999-06-04

NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 167
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
US-09-325-932A-167

Query Match 78.1%; Score 82; DB 4; Length 148;
Best Local Similarity 70.0%; Pred. No. 4.3e-06;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VPOSIDMRDGAATSVKNKG 20
Db 57 LPKDFMRDHGAVTAVVKNKG 76

RESULT 8
US-09-325-932A-165
; Sequence 165, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325.932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 165
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
US-09-325-932A-165

Query Match 78.1%; Score 82; DB 4; Length 278;
Best Local Similarity 70.0%; Pred. No. 9.1e-06;
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VPOSIDMRDGAATSVKNKG 20
Db 144 LPEDFMRDHGAVTAVVKNKG 163

RESULT 9
US-09-120-365-64
; Sequence 64, Application US/09120365
; Patent No. 6103514
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/120.365
; CURRENT FILING DATE: 1998-07-22
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Murine cathepsin L
US-09-120-365-64

Query Match 78.1%; Score 82; DB 3; Length 334;
Best Local Similarity 65.0%; Pred. No. 1.1e-05;
Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VPOSIDMRDGAATSVKNKG 20
Db 114 IPKSDMRKGVTPVKNKG 133

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Sarcophaga cathepsin L
US-09-120-365-80

Query Match      78.1%; Score 82; DB 3; Length 339;
Best Local Similarity 70.0%; Pred. No. 1.2e-05;
Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSCGAVTSVKNOG 20
   ||:||||:||||:||||:||||:
Db 122 VPKSVDWREHGAVTGKDOG 141

RESULT 13
US-09-515-039-62
; Sequence 62, Application US/09515039
; Patent No. 6214599
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/515,039
; CURRENT FILING DATE: 2000-03-06
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Sarcophaga cathepsin L
US-09-515-039-62

Query Match      78.1%; Score 82; DB 4; Length 339;
Best Local Similarity 70.0%; Pred. No. 1.2e-05;
Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSCGAVTSVKNOG 20
   ||:||||:||||:||||:||||:
Db 122 VPKSVDWREHGAVTGKDOG 141

RESULT 14
US-09-515-039-80
; Sequence 80, Application US/09515039
; Patent No. 6214599
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/515,039
; CURRENT FILING DATE: 2000-03-06
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Sarcophaga cathepsin L
US-09-515-039-80

Query Match      78.1%; Score 82; DB 4; Length 339;
Best Local Similarity 70.0%; Pred. No. 1.2e-05;
Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSCGAVTSVKNOG 20
   ||:||||:||||:||||:||||:
Db 122 VPKSVDWREHGAVTGKDOG 141

RESULT 12
US-09-120-365-80
; Sequence 80, Application US/09120365
; Patent No. 6103514
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/120,365
; CURRENT FILING DATE: 1998-07-22
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Sarcophaga cathepsin L
US-09-120-365-62

Query Match      78.1%; Score 82; DB 3; Length 339;
Best Local Similarity 70.0%; Pred. No. 1.2e-05;
Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSCGAVTSVKNOG 20
   ||:||||:||||:||||:||||:
Db 122 VPKSVDWREHGAVTGKDOG 141

RESULT 12
US-09-120-365-80
; Sequence 80, Application US/09120365
; Patent No. 6103514
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/120,365
; CURRENT FILING DATE: 1998-07-22
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Sarcophaga cathepsin L
US-09-120-365-62

Query Match      78.1%; Score 82; DB 4; Length 334;
Best Local Similarity 65.0%; Pred. No. 1.1e-05;
Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSCGAVTSVKNOG 20
   ||:||||:||||:||||:||||:
Db 114 IPKSVDWREKGCVTVPKNOG 133

RESULT 11
US-09-120-365-62
; Sequence 62, Application US/09120365
; Patent No. 6103514
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/120,365
; CURRENT FILING DATE: 1998-07-22
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Sarcophaga cathepsin L
US-09-120-365-62

Query Match      78.1%; Score 82; DB 4; Length 334;
Best Local Similarity 65.0%; Pred. No. 1.1e-05;
Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSCGAVTSVKNOG 20
   ||:||||:||||:||||:||||:
Db 114 IPKSVDWREKGCVTVPKNOG 133

RESULT 11
US-09-120-365-62
; Sequence 62, Application US/09120365
; Patent No. 6103514
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/120,365
; CURRENT FILING DATE: 1998-07-22
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Sarcophaga cathepsin L
US-09-120-365-62

Query Match      78.1%; Score 82; DB 3; Length 339;
Best Local Similarity 70.0%; Pred. No. 1.2e-05;
Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSCGAVTSVKNOG 20
   ||:||||:||||:||||:||||:
Db 122 VPKSVDWREHGAVTGKDOG 141

RESULT 12
US-09-120-365-80
; Sequence 80, Application US/09120365
; Patent No. 6103514
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/120,365
; CURRENT FILING DATE: 1998-07-22
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Sarcophaga cathepsin L
US-09-120-365-80
```


RESULT 15
US-09-325-932A-153

; Sequence 153, Application US/09325932A

; Patent No. 6451604

; GENERAL INFORMATION:

; APPLICANT: Film, Barry

; APPLICANT: Lasham, Annette

; TITLE OF INVENTION: Compositions affecting programmed cell

; TITLE OF INVENTION: death and their use in the modification of forestry plant devel

; FILE REFERENCE: 1022

; CURRENT APPLICATION NUMBER: US/09/325,932A

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 206

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 153

; LENGTH: 396

; TYPE: PRT

; ORGANISM: Pinus radiata

US-09-325-932A-153

Query Match 77.1%; Score 81; DB 4; Length 396;

Best Local Similarity 70.0%; Pred. No. 2e-05;

Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPSIDWRDGAATSVKNOG 20

Db 160 LPSIDWREKGAATSVKNOG 179

Search completed: November 21, 2002, 16:11:01
Job time : 11.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2002, 16:08:44 (Search time 137.5 Seconds
(Without alignments)
93.779 Million cell updates/sec

Title: US-09-674-738-1

Perfect score: 108

Sequence: 1 VPOSIDMRDYGAVTSVKNQ 20

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Pending_Patents_Aa Main:*

1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep:*

2: /cgn2_6/ptodata/1/paa/US06_COMB.pep:*

3: /cgn2_6/ptodata/1/paa/US07_COMB.pep:*

4: /cgn2_6/ptodata/1/paa/US08_COMB.pep:*

5: /cgn2_6/ptodata/1/paa/US081_COMB.pep:*

6: /cgn2_6/ptodata/1/paa/US082_COMB.pep:*

7: /cgn2_6/ptodata/1/paa/US083_COMB.pep:*

8: /cgn2_6/ptodata/1/paa/US084_COMB.pep:*

9: /cgn2_6/ptodata/1/paa/US085_COMB.pep:*

10: /cgn2_6/ptodata/1/paa/US086_COMB.pep:*

11: /cgn2_6/ptodata/1/paa/US087_COMB.pep:*

12: /cgn2_6/ptodata/1/paa/US088_COMB.pep:*

13: /cgn2_6/ptodata/1/paa/US089_COMB.pep:*

14: /cgn2_6/ptodata/1/paa/US090_COMB.pep:*

15: /cgn2_6/ptodata/1/paa/US091_COMB.pep:*

16: /cgn2_6/ptodata/1/paa/US092_COMB.pep:*

17: /cgn2_6/ptodata/1/paa/US093_COMB.pep:*

18: /cgn2_6/ptodata/1/paa/US094_COMB.pep:*

19: /cgn2_6/ptodata/1/paa/US095_COMB.pep:*

20: /cgn2_6/ptodata/1/paa/US096_COMB.pep:*

21: /cgn2_6/ptodata/1/paa/US097_COMB.pep:*

22: /cgn2_6/ptodata/1/paa/US098_COMB.pep:*

23: /cgn2_6/ptodata/1/paa/US099_COMB.pep:*

24: /cgn2_6/ptodata/1/paa/US100_COMB.pep:*

25: /cgn2_6/ptodata/1/paa/US101_COMB.pep:*

26: /cgn2_6/ptodata/1/paa/US102_COMB.pep:*

27: /cgn2_6/ptodata/1/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the distribution, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	108	100.0	20	US-09-674-738-1
2	108	100.0	21	US-09-382-685A-2
3	108	100.0	21	US-09-382-688-3
4	108	100.0	21	US-09-750-210A-3
5	108	100.0	212	US-08-244-796-15
6	108	100.0	356	US-09-791-537-84879

7	108	100.0	356	21	US-09-791-537-84882	Sequence 84882, A
8	99	91.7	311	21	US-09-743-987-16	Sequence 16, Appl
9	99	91.7	324	21	US-09-743-987-20	Sequence 20, Appl
10	99	91.7	324	21	US-09-791-537-69324	Sequence 69324, A
11	99	91.7	351	21	US-09-743-987-14	Sequence 14, Appl
12	99	91.7	351	21	US-09-791-537-84889	Sequence 84889, A
13	97	89.8	20	17	US-09-382-685A-4	Sequence 4, Appl
14	97	89.8	20	17	US-09-382-688-5	Sequence 5, Appl
15	97	89.8	20	21	US-09-750-210A-5	Sequence 5, Appl
16	95	88.0	186	21	US-09-743-987-2	Sequence 2, Appl
17	95	88.0	186	21	US-09-743-987-8	Sequence 8, Appl
18	95	88.0	187	21	US-09-743-987-4	Sequence 4, Appl
19	95	88.0	187	21	US-09-743-987-6	Sequence 6, Appl
20	95	88.0	190	21	US-09-743-987-18	Sequence 18, Appl
21	93	86.1	20	17	US-09-382-685A-3	Sequence 3, Appl
22	93	86.1	20	17	US-09-382-688-4	Sequence 4, Appl
23	93	86.1	20	20	US-09-674-738-2	Sequence 2, Appl
24	93	86.1	20	21	US-09-750-210A-4	Sequence 4, Appl
25	93	86.1	216	21	US-09-791-537-84208	Sequence 84208, A
26	93	86.1	345	21	US-09-791-537-84883	Sequence 84883, A
27	92	85.2	357	21	US-09-791-537-84884	Sequence 84884, A
28	90	83.3	312	21	US-09-791-537-110767	Sequence 110767, A
29	90	83.3	340	21	US-09-791-537-69330	Sequence 69330, A
30	88	81.5	18	17	US-09-382-688-2	Sequence 2, Appl
31	88	81.5	18	21	US-09-743-987-21	Sequence 21, Appl
32	88	81.5	18	21	US-09-750-210A-2	Sequence 21, Appl
33	88	81.5	18	24	US-10-029-307-1	Sequence 1, Appl
34	80	74.1	326	21	US-09-791-537-119891	Sequence 119891, A
35	80	74.1	454	21	US-09-791-537-34256	Sequence 34256, A
36	79	72.1	98	20	US-09-620-393B-3345	Sequence 3345, Ap
37	79	72.1	101	20	US-09-620-393B-3344	Sequence 3344, Ap
38	79	72.1	127	17	US-09-327-373-176	Sequence 176, App
39	79	72.1	127	23	US-09-990-004A-176	Sequence 176, App
40	79	72.1	127	26	US-10-219-220-167	Sequence 167, App
41	79	72.1	148	17	US-09-327-373-167	Sequence 167, App
42	79	72.1	148	23	US-09-990-004A-167	Sequence 167, App
43	79	72.1	148	26	US-10-219-220-167	Sequence 167, App
44	79	72.1	246	21	US-09-708-427-20930	Sequence 20930, A
45	79	72.1	258	19	US-09-513-996A-55464	Sequence 55464, A

ALIGNMENTS

RESULT 1

US-09-674-738-1

Sequence 1, Application US/09674738

GENERAL INFORMATION:

APPLICANT: UNSAPHAM Arzneimittel GmbH

TITLE OF INVENTION: Use of Bromelain Processes for Inhibiting Blood

FILE REFERENCE: 80054

CURRENT APPLICATION NUMBER: US/09/674,738

CURRENT FILING DATE: 2000-10-31

PRIOR APPLICATION NUMBER: PCT/EP98/04406

PRIOR FILING DATE: 1998-07-15

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 20

TYPE: PRT

ORGANISM: pine-apple (Bromeliaceae)

US-09-674-738-1

Query Match 100.0%; Score 108; DB 20;
Best Local Similarity 100.0%; Pred. No. 6; 9e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPOSIDMRDYGAVTSVKNQ 20
Db 1 VPOSIDMRDYGAVTSVKNQ 20

RESULT 2

US-09-382-685A-2
; Sequence 2, Application US/09382685A
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracy Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0750001
; CURRENT APPLICATION NUMBER: US/09/382,685A
; CURRENT FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00591
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; PRIOR FILING DATE: 1997-02-25
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 21
; TYPE: PRT
; ORGANISM: ORGANISM: Ananas comosus
US-09-382-685A-2

Query Match 100.0%; Score 108; DB 17; Length 21;
Best Local Similarity 100.0%; Pred. No. 7.3e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKQN 20
|||
DB 2 VPOSIDWRDYGAVTSVKQN 21
|||

RESULT 3

US-09-382-688-3
; Sequence 3, Application US/09382688A
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracy Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0740001
; CURRENT APPLICATION NUMBER: US/09/382,688A
; CURRENT FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00590
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; PRIOR FILING DATE: 1997-02-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 21
; TYPE: PRT
; ORGANISM: ORGANISM: Ananas comosus
US-09-382-688-3

Query Match 100.0%; Score 108; DB 17; Length 21;
Best Local Similarity 100.0%; Pred. No. 7.3e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKQN 20
|||

Db

2 VPOSIDWRDYGAVTSVKQN 21

RESULT 4

US-09-750-210A-3
; Sequence 3, Application US/09750210A
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracy Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0740001
; CURRENT APPLICATION NUMBER: US/09/750,210A
; CURRENT FILING DATE: 2000-12-29
; PRIOR APPLICATION NUMBER: US 09/382,688
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00590
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; PRIOR FILING DATE: 1997-02-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 21
; TYPE: PRT
; ORGANISM: ORGANISM: Ananas comosus
US-09-750-210A-3

Query Match 100.0%; Score 108; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 7.3e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKQN 20
|||
DB 2 VPOSIDWRDYGAVTSVKQN 21
|||

RESULT 5

US-08-244-796-15
; Sequence 15, Application US/08244796
; GENERAL INFORMATION:
; APPLICANT: ANDREW R. MILNER, MICHAEL PANACCIO, TERENCE W.
; TITLE OF INVENTION: LIVER FLUKE VACCINE AND POLYPEPTIDES USEFUL FOR SAME
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER
; STREET: 400 GARDEN CITY PLAZA, GARDEN CITY
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 11530-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/244,796
; FILING DATE: 15-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PL7109
; FILING DATE: 05-FEB-1993
; APPLICATION NUMBER: PCT/AU94/00051
; FILING DATE: 04-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: DIGIGLIO, FRANK
; REGISTRATION NUMBER: 31,346

REFERENCE/DOCKET NUMBER: 9316
TELECOMMUNICATION INFORMATION:
TELEPHONE: (1516) 742 4343
TELEFAX: (1516) 742 4366
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 212 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-244-796-15

Query Match 100.0%; Score 108; DB 6; Length 212;
Best Local Similarity 100.0%; Pred. No. 9.1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPOSIDMRDYGAVTSVKNN 20
Db 2 VPOSIDMRDYGAVTSVKNN 21

RESULT 6
US-09-791-537-84879
; Sequence 84879, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMEB
; FILE REFERENCE: 261/210
; CURRENT FILING DATE: 2001-02-22
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 84879
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Ananas comosus
US-09-791-537-84879

Query Match 100.0%; Score 108; DB 21; Length 356;
Best Local Similarity 100.0%; Pred. No. 1.6e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPOSIDMRDYGAVTSVKNN 20
Db 124 VPOSIDMRDYGAVTSVKNN 143

RESULT 7
US-09-791-537-84882
; Sequence 84882, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danner, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMEB
; FILE REFERENCE: 261/210
; CURRENT FILING DATE: 2001-02-22
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 84882
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Ananas comosus
US-09-791-537-84882

Query Match 100.0%; Score 108; DB 21; Length 356;
Best Local Similarity 100.0%; Pred. No. 1.6e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 VPOSIDMRDYGAVTSVKNN 20
Db 124 VPOSIDMRDYGAVTSVKNN 143

RESULT 8
US-09-743-987-16
; Sequence 16, Application US/09743987
; GENERAL INFORMATION:
; APPLICANT: Wynnott, Tracey Lehane
; APPLICANT: Crosssett, Ben
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM PINEAPPLE STEM
; FILE REFERENCE: 0623.0980000
; CURRENT FILING DATE: 2001-08-01
; CURRENT FILING DATE: 2001-08-01
; PRIOR FILING DATE: 1999-08-24
; PRIOR FILING DATE: 1998-09-02
; PRIOR FILING DATE: 1998-09-02
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Ananas comosus
US-09-743-987-16

Query Match 91.7%; Score 99; DB 21; Length 311;
Best Local Similarity 90.0%; Pred. No. 4e-07;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VPOSIDMRDYGAVTSVKNN 20
Db 96 VPOSIDMRDYGAVTSVKNN 115

RESULT 9
US-09-743-987-20
; Sequence 20, Application US/09743987
; GENERAL INFORMATION:
; APPLICANT: Wynnott, Tracey Lehane
; APPLICANT: Crosssett, Ben
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM PINEAPPLE STEM
; FILE REFERENCE: 0623.0980000
; CURRENT FILING DATE: 2001-08-01
; CURRENT FILING DATE: 2001-08-01
; PRIOR FILING DATE: 1999-08-24
; PRIOR FILING DATE: 1998-09-02
; PRIOR FILING DATE: 1998-09-02
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Ananas comosus
US-09-743-987-20

Query Match 91.7%; Score 99; DB 21; Length 324;
Best Local Similarity 90.0%; Pred. No. 4.1e-07;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VPOSIDMRDYGAVTSVKNN 20
Db 96 VPOSIDMRDYGAVTSVKNN 115

RESULT 10
US-09-791-537-69324
; Sequence 69324, Application US/09791537
; GENERAL INFORMATION:

```
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBE
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791.537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 69324
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Ananas comosus
US-09-791-537-69324

Query Match      91.7%; Score 99; DB 21; Length 324;
Best Local Similarity 90.0%; Pred. No. 4.1e-07;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VPOSIDWRDYGAVTSVKQN 20
Db 96 VPOSIDWRDYGAVNEVKQN 115

RESULT 11
US-09-743-987-14
; Sequence 14, Application US/09743987
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracey Lehanne
; APPLICANT: Crossett, Ben
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM PINEAPPLE STEM
; FILE REFERENCE: 0623.0980000
; CURRENT APPLICATION NUMBER: US/09/743.987
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: PCT/GB99/02784
; PRIOR FILING DATE: 1999-08-24
; PRIOR APPLICATION NUMBER: GB 9819138
; PRIOR FILING DATE: 1998-09-02
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Ananas comosus
US-09-743-987-14

Query Match      91.7%; Score 99; DB 21; Length 351;
Best Local Similarity 90.0%; Pred. No. 4.5e-07;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VPOSIDWRDYGAVTSVKQN 20
Db 123 VPOSIDWRDYGAVNEVKQN 142

RESULT 12
US-09-791-537-84889
; Sequence 84889, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBE
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791.537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 84889
; LENGTH: 351
; TYPE: PRT
```

```
; ORGANISM: Ananas comosus
US-09-791-537-84889

Query Match      91.7%; Score 99; DB 21; Length 351;
Best Local Similarity 90.0%; Pred. No. 4.5e-07;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VPOSIDWRDYGAVTSVKQN 20
Db 123 VPOSIDWRDYGAVNEVKQN 142

RESULT 13
US-09-382-685A-4
; Sequence 4, Application US/09382685A
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracey Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0750001
; CURRENT APPLICATION NUMBER: US/09/382.685A
; CURRENT FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00591
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Ananas comosus
US-09-382-685A-4

Query Match      89.8%; Score 97; DB 17; Length 20;
Best Local Similarity 94.7%; Pred. No. 4.2e-08;
Matches 18; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPOSIDWRDYGAVTSVKNQ 19
Db 1 VPOSIDWRNIGAVTSVKNQ 19

RESULT 14
US-09-382-688-5
; Sequence 5, Application US/09382688A
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracey Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0740001
; CURRENT APPLICATION NUMBER: US/09/382.688A
; CURRENT FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00590
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
```

SEQ ID NO 5
LENGTH: 20
TYPE: PRT
ORGANISM: Ananas comosus
US-09-382-688-5

Query Match 89.8%; Score 97; DB 17; Length 20;
Best Local Similarity 94.7%; Pred. No. 4.2e-08;
Matches 18; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPOSIDMRDYGAVTYKNO 19
|||||:|||||
Db 1 VPOSIDMRNYGAVTYKNO 19

RESULT 15

US-09-750-210A-5
Sequence 5, Application US/09750210A
GENERAL INFORMATION:
APPLICANT: Mynott, Tracy Leanne
APPLICANT: Engwerda, Christian
APPLICANT: Peek, Keith
TITLE OF INVENTION: Component of Bromelain
FILE REFERENCE: 0623.074001
CURRENT APPLICATION NUMBER: US/09/750,210A
CURRENT FILING DATE: 2000-12-29
PRIOR APPLICATION NUMBER: US 09/382,688
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: PCT/GB98/00590
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: GB 9706119.6
PRIOR FILING DATE: 1997-03-25
PRIOR APPLICATION NUMBER: GB 9704252.7
PRIOR FILING DATE: 1997-02-28
PRIOR APPLICATION NUMBER: GB 9703850.9
PRIOR FILING DATE: 1997-02-25
PRIOR APPLICATION NUMBER: GB 9703827.7
PRIOR FILING DATE: 1997-02-25
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 20
TYPE: PRT
ORGANISM: ORGANISM: Ananas comosus
US-09-750-210A-5

Query Match 89.8%; Score 97; DB 21; Length 20;
Best Local Similarity 94.7%; Pred. No. 4.2e-08;
Matches 18; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPOSIDMRDYGAVTYKNO 19
|||||:|||||
Db 1 VPOSIDMRNYGAVTYKNO 19

Search completed: November 21, 2002, 16:15:42
Job time : 138.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2002, 16:08:44 ; Search time 137.5 Seconds
(without alignments)

93.779 Million cell updates/sec

Title: US-09-674-738-2

Perfect score: 105

Sequence: 1 VPSIDWRDSCAVTSVKNQG 20

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

Pending Patents_AA_Main:*

1: /cgn2_6/ptodata/1/paa/PCRTUS_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/US097_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US099_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
25: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
26: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*
27: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by a result to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	105	100.0	20 17	US-09-382-685A-3
2	105	100.0	20 17	US-09-382-688-4
3	105	100.0	20 20	US-09-674-738-2
4	105	100.0	20 21	US-09-750-210A-4
5	105	100.0	216 21	US-09-791-537-82408
6	105	100.0	345 21	US-09-791-537-84883

7	94	89.5	20 17	US-09-382-685A-4	Sequence 4, Appli
8	94	89.5	20 17	US-09-382-688-5	Sequence 5, Appli
9	94	89.5	20 21	US-09-750-210A-5	Sequence 5, Appli
10	93	88.6	20 20	US-09-674-738-1	Sequence 1, Appli
11	93	88.6	21 17	US-09-382-685A-2	Sequence 2, Appli
12	93	88.6	21 17	US-09-382-688-3	Sequence 3, Appli
13	93	88.6	21 21	US-09-750-210A-3	Sequence 3, Appli
14	93	88.6	212 6	US-08-244-796-15	Sequence 15, Appli
15	93	88.6	356 21	US-09-791-537-84879	Sequence 84879, A
16	93	88.6	356 21	US-09-791-537-84882	Sequence 84882, A
17	88	83.8	454 21	US-09-791-537-82256	Sequence 32256, A
18	85	81.0	334 21	US-09-791-537-84711	Sequence 84711, A
19	84	80.0	311 21	US-09-743-987-16	Sequence 16, Appli
20	84	80.0	324 21	US-09-743-987-20	Sequence 20, Appli
21	84	80.0	351 21	US-09-791-537-69324	Sequence 69324, A
22	84	80.0	351 21	US-09-743-987-14	Sequence 14, Appli
23	84	80.0	351 21	US-09-791-537-84889	Sequence 84889, A
24	83	79.0	326 26	US-10-219-999-45992	Sequence 45992, A
25	83	79.0	334 19	US-09-519-283-4	Sequence 4, Appli
26	83	79.0	334 21	US-09-791-537-82405	Sequence 82405, A
27	83	79.0	334 25	US-10-143-650-19	Sequence 19, Appli
28	83	79.0	357 21	US-09-791-537-84884	Sequence 84884, A
29	83	79.0	374 21	US-09-791-537-91358	Sequence 91358, A
30	83	79.0	399 26	US-10-219-999-60863	Sequence 60863, A
31	83	79.0	399 27	US-09-324-109-29193	Sequence 29193, A
32	83	79.0	443 19	US-09-515-038-69	Sequence 69, Appli
33	83	79.0	443 21	US-09-791-537-56752	Sequence 56752, A
34	83	79.0	443 21	US-09-791-537-68257	Sequence 68257, A
35	83	79.0	498 21	US-09-791-537-34918	Sequence 34918, A
36	83	79.0	498 27	US-09-360-039-6416	Sequence 6416, Ap
37	82	78.1	127 13	US-09-327-373-176	Sequence 176, App
38	82	78.1	127 23	US-09-990-004-176	Sequence 176, App
39	82	78.1	127 26	US-10-219-220-176	Sequence 176, App
40	82	78.1	148 17	US-09-327-373-167	Sequence 167, App
41	82	78.1	148 25	US-09-390-004-167	Sequence 167, App
42	82	78.1	148 26	US-10-219-220-167	Sequence 167, App
43	82	78.1	212 21	US-09-791-537-13192	Sequence 13192, A
44	82	78.1	278 17	US-09-327-373-165	Sequence 165, App
45	82	78.1	278 23	US-09-990-004-165	Sequence 165, App

ALIGNMENTS

RESULT 1
US-09-382-685A-3
: GENERAL INFORMATION:
: APPLICANT: Mynott, Tracy Lehane
: APPLICANT: Engwerda, Christian
: APPLICANT: Peek, Keith
: TITLE OF INVENTION: Component of Bromelain
: FILE REFERENCE: 0623.075001
: CURRENT APPLICATION NUMBER: US/09/382, 685A
: PRIOR FILING DATE: 1999-08-25
: PRIOR APPLICATION NUMBER: PCT/GB98/00591
: PRIOR FILING DATE: 1998-02-25
: PRIOR APPLICATION NUMBER: GB 9706119.6
: PRIOR FILING DATE: 1997-03-25
: PRIOR APPLICATION NUMBER: GB 9704252.7
: PRIOR FILING DATE: 1997-02-28
: PRIOR APPLICATION NUMBER: GB 9703850.9
: PRIOR FILING DATE: 1997-02-25
: PRIOR APPLICATION NUMBER: GB 9703827.7
: PRIOR FILING DATE: 1997-02-25
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 3
: LENGTH: 20
: TYPE: PRT
: ORGANISM: ORGANISM: Ananas comosus
US-09-382-685A-3

Query Match 100.0%; Score 105; DB 17; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSGAVTSVKNOG 20
|||
Db 1 VPOSIDWRDSGAVTSVKNOG 20

RESULT 2

US-09-382-688-4
; Sequence 4, Application US/09382688A
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracy Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0740001
; CURRENT APPLICATION NUMBER: US/09/382.688A
; CURRENT FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00590
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; PRIOR FILING DATE: 1997-02-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Ananas comosus
US-09-382-688-4

Query Match 100.0%; Score 105; DB 17; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSGAVTSVKNOG 20
|||
Db 1 VPOSIDWRDSGAVTSVKNOG 20

RESULT 3

US-09-674-738-2
; Sequence 2, Application US/09674738
; GENERAL INFORMATION:
; APPLICANT: URSAPHARM Arzneimittel GmbH
; TITLE OF INVENTION: Use of Bromelain Proteases for Inhibiting Blood
; FILE REFERENCE: 80054
; CURRENT APPLICATION NUMBER: US/09/674.738
; CURRENT FILING DATE: 2000-10-31
; PRIOR APPLICATION NUMBER: PCT/EP98/04406
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 20
; TYPE: PRT
; ORGANISM: pine-apple (Bromeliaceae)
US-09-674-738-2

Query Match 100.0%; Score 105; DB 20; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSGAVTSVKNOG 20
|||

Db 1 VPOSIDWRDSGAVTSVKNOG 20

RESULT 4

US-09-750-210A-4
; Sequence 4, Application US/09750210A
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracy Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0740001
; CURRENT APPLICATION NUMBER: US/09/750.210A
; CURRENT FILING DATE: 2000-12-29
; PRIOR APPLICATION NUMBER: US 09/382,688
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00590
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; PRIOR FILING DATE: 1997-02-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Ananas comosus
US-09-750-210A-4

Query Match 100.0%; Score 105; DB 21; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSGAVTSVKNOG 20
|||
Db 1 VPOSIDWRDSGAVTSVKNOG 20

RESULT 5

US-09-791-82408
; Sequence 82408, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBER
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 82408
; LENGTH: 216
; TYPE: PRT
; ORGANISM: Ananas comosus
US-09-791-82408

Query Match 100.0%; Score 105; DB 21; Length 216;
Best Local Similarity 100.0%; Pred. No. 4.9e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSGAVTSVKNOG 20
|||
Db 1 VPOSIDWRDSGAVTSVKNOG 20

RESULT 6

```
US-09-791-537-84883
; Sequence 84883, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Biocomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBE
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791, 537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 84883
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Ananas comosus
US-09-791-537-84883

Query Match          100.0%; Score 105; DB 21; Length 345;
Best Local Similarity 100.0%; Pred. No. 8.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  VPOSIDWRDSCGAVTSVKNOG 20
Db      123 VPOSIDWRDSCGAVTSVKNOG 142

RESULT 7
US-09-382-685A-4
; Sequence 4, Application US/09382685A
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracy Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0750001
; CURRENT APPLICATION NUMBER: US/09/382, 685A
; CURRENT FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00591
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28 9703850.9
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; PRIOR FILING DATE: 1997-02-25
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 20
; TYPE: PRT
; ORGANISM: ORGANISM: Ananas comosus
US-09-382-685A-4

Query Match          89.5%; Score 94; DB 17; Length 20;
Best Local Similarity 90.0%; Pred. No. 2.1e-07;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1  VPOSIDWRDSCGAVTSVKNOG 20
Db      1  VPOSIDWRDSCGAVTSVKNOG 20

RESULT 8
US-09-382-688-5
; Sequence 5, Application US/09382688A
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracy Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
```

```
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0740001
; CURRENT APPLICATION NUMBER: US/09/382, 688A
; CURRENT FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00590
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28 9703850.9
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; PRIOR FILING DATE: 1997-02-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 20
; TYPE: PRT
; ORGANISM: ORGANISM: Ananas comosus
US-09-750-210A-5

Query Match          89.5%; Score 94; DB 17; Length 20;
Best Local Similarity 90.0%; Pred. No. 2.1e-07;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1  VPOSIDWRDSCGAVTSVKNOG 20
Db      1  VPOSIDWRDSCGAVTSVKNOG 20

RESULT 9
US-09-750-210A-5
; Sequence 5, Application US/09750210A
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracy Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0740001
; CURRENT APPLICATION NUMBER: US/09/750, 210A
; CURRENT FILING DATE: 2000-12-29
; PRIOR APPLICATION NUMBER: US 09/382, 688
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00590
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; PRIOR FILING DATE: 1997-02-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 20
; TYPE: PRT
; ORGANISM: ORGANISM: Ananas comosus
US-09-750-210A-5

Query Match          89.5%; Score 94; DB 21; Length 20;
Best Local Similarity 90.0%; Pred. No. 2.1e-07;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1  VPOSIDWRDSCGAVTSVKNOG 20
Db      1  VPOSIDWRDSCGAVTSVKNOG 20

RESULT 10
US-09-674-738-1
```

```

; Sequence 1, Application US/09674738
; GENERAL INFORMATION:
; APPLICANT: URSPHARM Arzneimittel GmbH
; TITLE OF INVENTION: Use of Bromelain Proteases for Inhibiting Blood
; TITLE OF INVENTION: Coagulation
; FILE REFERENCE: 80054
; CURRENT APPLICATION NUMBER: US/09/674,738
; PRIOR FILING DATE: 2000-10-31
; PRIOR APPLICATION NUMBER: PCT/EP98/04406
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 20
; TYPE: PRT
; ORGANISM: pine-apple (Bromeliaceae)
US-09-674-738-1

Query Match      88.6%; Score 93; DB 20; Length 20;
Best Local Similarity 94.7%; Pred. No. 3.1e-07;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKNQ 19
| | | | | | | | | | | | | | | | | |
DB 1 VPOSIDWRDYGAVTSVKNQ 19

RESULT 11
US-09-382-685A-2
; Sequence 2, Application US/09382685A
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracy Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0750001
; CURRENT APPLICATION NUMBER: US/09/382,685A
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00591
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 21
; TYPE: PRT
; ORGANISM: ORGANISM: Ananas comosus
US-09-382-685A-2

Query Match      88.6%; Score 93; DB 17; Length 21;
Best Local Similarity 94.7%; Pred. No. 3.2e-07;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKNQ 19
| | | | | | | | | | | | | | | | | |
DB 2 VPOSIDWRDYGAVTSVKNQ 20

RESULT 12
US-09-382-688-3
; Sequence 3, Application US/09382688A
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracy Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0740001
; CURRENT APPLICATION NUMBER: US/09/750,210A
; PRIOR FILING DATE: 2000-12-29
; PRIOR APPLICATION NUMBER: US 09/382,688
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00590
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 21
; TYPE: PRT
; ORGANISM: ORGANISM: Ananas comosus
US-09-750-210A-3

Query Match      88.6%; Score 93; DB 21; Length 21;
Best Local Similarity 94.7%; Pred. No. 3.2e-07;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKNQ 19
| | | | | | | | | | | | | | | | | |
DB 2 VPOSIDWRDYGAVTSVKNQ 20

RESULT 14
US-08-244-796-15
; Sequence 15, Application US/08244796

```

```

; FILE REFERENCE: 0623.0740001
; CURRENT APPLICATION NUMBER: US/09/382,688A
; CURRENT FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00590
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 21
; TYPE: PRT
; ORGANISM: ORGANISM: Ananas comosus
US-09-382-688-3

Query Match      88.6%; Score 93; DB 17; Length 21;
Best Local Similarity 94.7%; Pred. No. 3.2e-07;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKNQ 19
| | | | | | | | | | | | | | | | | |
DB 2 VPOSIDWRDYGAVTSVKNQ 20

RESULT 13
US-09-750-210A-3
; Sequence 3, Application US/09750210A
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracy Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0740001
; CURRENT APPLICATION NUMBER: US/09/750,210A
; CURRENT FILING DATE: 2000-12-29
; PRIOR APPLICATION NUMBER: US 09/382,688
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00590
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 21
; TYPE: PRT
; ORGANISM: ORGANISM: Ananas comosus
US-09-750-210A-3

Query Match      88.6%; Score 93; DB 21; Length 21;
Best Local Similarity 94.7%; Pred. No. 3.2e-07;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKNQ 19
| | | | | | | | | | | | | | | | | |
DB 2 VPOSIDWRDYGAVTSVKNQ 20

RESULT 14
US-08-244-796-15
; Sequence 15, Application US/08244796

```

GENERAL INFORMATION:
 APPLICANT: ANDREW R. MILNER, MICHAEL PANACCIO, TERENCE W.
 TITLE OF INVENTION: LIVER FLUKE VACCINE AND POLYPEPTIDES USEFUL FOR SAME
 NUMBER OF SEQUENCES: 24
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER
 STREET: 400 GARDEN CITY PLAZA, GARDEN CITY
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: UNITED STATES OF AMERICA
 ZIP: 11530-0299
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentln Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US//08/244,796
 FILING DATE: 15-AUG-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: AU PL7109
 FILING DATE: 05-FEB-1993
 APPLICATION NUMBER: PCT/AU94/00051
 FILING DATE: 04-FEB-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: DIGILIO, FRANK
 REGISTRATION NUMBER: 31,346
 REFERENCE/DOCKET NUMBER: 9116
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (1516) 742 4343
 TELEFAX: (1516) 742 4366
 INFORMATION FOR SEQ ID NO: 15:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 212 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-244-796-15

Query Match 88.6%; Score 93; DB 6; Length 212;
 Best Local Similarity 94.7%; Pred. No. 4e-06;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKNO 19
 |||||
 Db 2 VPOSIDWRDYGAVTSVKNO 20

RESULT 15
 US-09-791-537-84879
 GENERAL INFORMATION:
 APPLICANT: Bionomix, Inc.
 APPLICANT: Debe, Derek
 APPLICANT: Danzer, Joseph
 TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
 FILE REFERENCE: 261/210
 CURRENT APPLICATION NUMBER: US/09/791,537
 FILING DATE: 2001-02-22
 NUMBER OF SEQ ID NOS: 153055
 SOFTWARE: Patentln version 3.0
 SEQ ID NO 84879
 LENGTH: 356
 TYPE: PRT
 ORGANISM: Ananas comosus
 US-09-791-537-84879

Query Match 88.6%; Score 93; DB 21; Length 356;
 Best Local Similarity 94.7%; Pred. No. 7e-06;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKNO 19
 |||||
 Db 124 VPOSIDWRDYGAVTSVKNO 142
 Search completed: November 21, 2002, 16:15:43
 Job time : 138.5 secs


```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-80627

```

Query Match	64.8%;	Score 70;	DB 5;	Length 230;
Best Local Similarity	63.2%;	Pred. No. 0.00045;		
Matches 12;	Conservative 4;	Mismatches 3;	Indels	

```
QY      1 VPQSIDWRDYGAVTSVKNQ 19
      :|:|:|:|:| | | | | |
Db      115 LPRSDWREKGYVTPVKNQ 133
```

```

US-09-724-676-80628
US-09-724-676-80628
; Sequence 80628, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen Ltd
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; NUMBER OF SEQ ID NOS: 91222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 80628
; LENGTH: 230
; TYPE: PRM
; ORGANISM: Homo sapiens
US-09-724-676-80628

```

Query Match	64.8%	Score 70	DB 5	Length 230
Beet Local Similarity	63.2%	Pred. No. 0.00045		
Matches 12; Conservative	4	Mismatches 3	Indels 0	Gaps 0

```
QY      1 VPQSIDWRDYGAVTSVKNQ 19  
        :|:||||:||||||  
Db     115 LPRSDWREKGYVTPVKNQ 133
```

```

RESULT 10
US-09-724-676-80629
: Sequence 80629, Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: Comugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129481.4 Comugen
: CURRENT APPLICATION NUMBER: US/09/724,676
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 80629
: LENGTH: 230
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-724-676-80629

```

Query Match	64.8%	Score 70;	DB 5;	Length 230;
Best Local Similarity	63.2%	Pred. No. 0.00045;		
Matches 12;	Conservative	4;	Mismatches 3;	Indels 0;
				Gaps 0

```
QY      1 VPQSIDWRDYGAVTSVKNQ 19  
          :|:|:|:|:|:|:|:|  
Db     115 LPRSVDWREKGYVTPVKNQ 133
```

```

; RESULT 11
; US-09-724-676-80630
; Sequence 80630, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676

```

```

;
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
;
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 80630
;
; LENGTH: 230
;
; TYPE: PRT
;
; ORGANISM: Homo sapiens
;
US-09-724-676-80630

```

Query Match 64.8%; Score 70; DB 5; Length 230;
Best Local Similarity 63.2%; Pred. No. 0.00045;
Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```
QY      1 VPOSIDWRDYGAVTSVKNO 19
        :|:|:|:|:| | | | |
Db      115 LPRSDVDRKREKYVTPVKNO 133
```

```

RESULT 12
US-09-724-676-80631
: Sequence 80631. Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: Comugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Comugen
: CURRENT APPLICATION NUMBER: US/09/724,676
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 80631
: LENGTH: 230
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-724-676-80631

```

Query Match	64.8%;	Score 70;	DB 5;	Length 230;
Best Local Similarity	63.2%;	Pred. No. 0.00045;		
Matches 12;	Conservative 4;	Mismatches 3;	Indels 0;	Gaps 0;

```
QY 1 VPQSIDWRDYGAVTSVKNO 19
    :|:|:|:|:|:|:|:|:|:|
Db 115 LPRSDWREKGYVTPVKNO 133
```

```

RESULT 13
US-09-724-676-80632
; Sequence 80632: Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ. ID NOS.: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 80632
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-80632

```

Query Match	64.8%	Score 70	DB 5	Length 230
Best Local Similarity	63.2%	Pred. No. 0.00045		
Matches 12	Conservative 4	Mismatches 3	Indels 0	Gaps 0

```
QY      1 VPQSIDWRDYGAVTSVKNQ 19
      :|:|:|:|:|:|:|:|:|:|
Db      115 LPRSDVWREKGYVTPVKNQ 133
```

RESULT 14
US-09-724-676-80588
: Sequence 80588, Application US/09724676

; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 80588
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-80588

Query Match 64.8%; Score 70; DB 5; Length 250;
Best Local Similarity 63.2%; Pred. No. 0.00049;
Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
Qy 1 VPOSIDWRDYGAVTSVKNQ 19
Db 115 LPRSDVDRKGYVTPVKQ 133

RESULT 15
US-09-724-676-80589
; Sequence 80589, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 80589
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-80589

Query Match 64.8%; Score 70; DB 5; Length 250;
Best Local Similarity 63.2%; Pred. No. 0.00049;
Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
Qy 1 VPOSIDWRDYGAVTSVKNQ 19
Db 115 LPRSDVDRKGYVTPVKQ 133

Search completed: November 21, 2002, 16:16:04
Job time : 8 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2002, 16:10:04 ; Search time 7 Seconds
(without alignments)
44.747 Million cell updates/sec

Title: US-09-674-738-1
Perfect score: 108
Sequence: 1 VPQSIDMRDYGAVTSVXNQN 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100480 seqs, 15661496 residues

Total number of hits satisfying chosen parameters: 100480

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: Published Applications_AA*
2: /cgn2_6/prodata/2/pubppaa/US08_NEW_PUB_PEP*
3: /cgn2_6/prodata/2/pubppaa/PCT_NEW_PUB_PEP*
4: /cgn2_6/prodata/2/pubppaa/US06_NEW_PUB_PEP*
5: /cgn2_6/prodata/2/pubppaa/US07_NEW_PUB_PEP*
6: /cgn2_6/prodata/2/pubppaa/US07_PUBCOMB_PEP*
7: /cgn2_6/prodata/2/pubppaa/PCTUS_PUBCOMB_PEP*
8: /cgn2_6/prodata/2/pubppaa/US08_PUBCOMB_PEP*
9: /cgn2_6/prodata/2/pubppaa/US09_NEW_PUB_PEP*
10: /cgn2_6/prodata/2/pubppaa/US10_PUBCOMB_PEP*
11: /cgn2_6/prodata/2/pubppaa/US10_NEW_PUB_PEP*
12: /cgn2_6/prodata/2/pubppaa/US10_PUBCOMB_PEP*
13: /cgn2_6/prodata/2/pubppaa/US60_NEW_PUB_PEP*
14: /cgn2_6/prodata/2/pubppaa/US60_PUBCOMB_PEP*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	100.0	21	10	US-09-750-210A-3
2	97	89.8	20	10	US-09-750-210A-5
3	93	86.1	20	10	US-09-750-210A-4
4	88	81.5	18	10	US-09-750-210A-2
5	71	65.7	21	10	US-09-750-210A-1
6	71	65.7	22	10	US-09-462-846-3
7	69	63.9	33	10	US-09-953-956-9
8	69	63.9	33	12	US-10-114-464-9
9	67	62.0	33	9	US-09-529-063-1
10	64	59.3	33	9	US-09-990-064-2
11	64	59.3	33	10	US-09-953-956-8
12	64	59.3	33	12	US-10-114-464-8
13	57	52.8	32	10	US-09-953-956-7
14	57	52.8	32	10	US-10-114-464-7
15	56	51.9	32	10	US-09-953-956-2
16	56	51.9	32	12	US-10-114-464-2
17	56	51.9	36	10	US-09-925-300-1693
18	55	50.9	33	10	US-09-925-301-1021
19	49.5	45.8	33	10	US-09-953-956-10

20	49.5	45.8	335	12	US-10-114-464-10	Sequence 10, Appl
21	49.5	45.8	364	10	US-09-925-302-478	Sequence 478, Ap
22	42	38.9	26	10	US-09-864-761-40428	Sequence 40428, A
23	42	38.9	425	9	US-09-486-734A-14	Sequence 14, Appl
24	41	38.0	185	10	US-09-893-737-78	Sequence 78, Appl
25	41	38.0	224	10	US-09-925-301-1302	Sequence 1302, Ap
26	41	38.0	565	10	US-09-922-217-1117	Sequence 1117, Ap
27	40.5	37.5	806	10	US-09-922-217-1118	Sequence 1118, Ap
28	40.5	37.5	806	10	US-10-052-586-454	Sequence 454, App
29	40.5	37.5	837	12	US-09-864-761-39892	Sequence 39892, A
30	40	37.0	68	10	US-09-925-301-1031	Sequence 1031, Ap
31	40	37.0	571	10	US-09-815-242-13686	Sequence 13686, A
32	40	37.0	637	10	US-09-862-007-2	Sequence 2, Appl1
33	40	37.0	1168	10	US-09-919-603-2	Sequence 2, Appl1
34	40	37.0	1172	10	US-09-919-770-4	Sequence 4, Appl1
35	40	37.0	1172	10	US-09-822-682-2	Sequence 2, Appl1
36	40	37.0	1247	10	US-09-815-242-10145	Sequence 10145, A
37	40	37.0	1247	10	US-09-815-242-13841	Sequence 13841, A
38	40	37.0	1247	10	US-09-864-761-42873	Sequence 42873, A
39	39.5	36.6	165	10	US-09-815-242-12007	Sequence 12007, A
40	39	36.1	394	10	US-09-788-626-20	Sequence 20, Appl
41	38	35.2	307	10	US-09-815-242-5135	Sequence 5135, Ap
42	38	35.2	408	10	US-09-815-242-11260	Sequence 11260, A
43	38	35.2	420	10	US-09-815-242-10370	Sequence 10370, A
44	38	35.2	457	10	US-10-001-852-12	Sequence 12, Appl
45	36	35.2	518	12		

ALIGNMENTS

RESULT 1
US-09-750-210A-3
Sequence 3, Application US/09750210A
Patent No. US20020102253A1
GENERAL INFORMATION:
APPLICANT: Mynott, Tracy lehanne
APPLICANT: Engwerda, Christian
TITLE OF INVENTION: Component of Bromelain
FILE REFERENCE: 0623.0740001
CURRENT APPLICATION NUMBER: US/09/750,210A
CURRENT FILING DATE: 2000-12-29
PRIOR APPLICATION NUMBER: US 09/382,688
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: PCT/GB98/00590
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: GB 9706119.6
PRIOR FILING DATE: 1997-03-25
PRIOR APPLICATION NUMBER: GB 9704252.7
PRIOR FILING DATE: 1997-02-28
PRIOR APPLICATION NUMBER: GB 9703850.9
PRIOR FILING DATE: 1997-02-25
PRIOR APPLICATION NUMBER: GB 9703827.7
PRIOR FILING DATE: 1997-02-25
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 21
TYPE: PRT
ORGANISM: Ananas comosus
US-09-750-210A-3

Query Match 100.0% Score 108; DB 10; Length 21;
Best Local Similarity 100.0%; Pred. No. 6.9e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VPQSIDMRDYGAVTSVXNQN 20
DB 2 VPQSIDMRDYGAVTSVXNQN 21
RESULT 2

```
US-09-750-210A-5
; Sequence 5, Application US/09750210A
; Patent No. US20020102253A1
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracy Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0740001
; CURRENT APPLICATION NUMBER: US/09/750,210A
; CURRENT FILING DATE: 2000-12-29
; PRIOR APPLICATION NUMBER: US 09/382,688
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00590
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; PRIOR FILING DATE: 1997-02-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 20
; TYPE: PRT
; ORGANISM: ORGANISM: Ananas comosus
US-09-750-210A-5

Query Match      89.8%; Score 97; DB 10; Length 20;
Best Local Similarity 94.7%; Pred. No. 4.4e-10;
Matches 18; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKNQ 19
DB 1 VPOSIDWRDYGAVTSVKNQ 19

RESULT 3
US-09-750-210A-4
; Sequence 4, Application US/09750210A
; Patent No. US20020102253A1
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracy Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0740001
; CURRENT APPLICATION NUMBER: US/09/750,210A
; CURRENT FILING DATE: 2000-12-29
; PRIOR APPLICATION NUMBER: US 09/382,688
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00590
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; PRIOR FILING DATE: 1997-02-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 20
; TYPE: PRT
; ORGANISM: ORGANISM: Ananas comosus
US-09-750-210A-4

Query Match      86.1%; Score 93; DB 10; Length 20;
Best Local Similarity 94.7%; Pred. No. 4.4e-10;
Matches 18; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKNQ 19
DB 1 VPOSIDWRDYGAVTSVKNQ 19

RESULT 4
US-09-750-210A-2
; Sequence 2, Application US/09750210A
; Patent No. US20020102253A1
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracy Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0740001
; CURRENT APPLICATION NUMBER: US/09/750,210A
; CURRENT FILING DATE: 2000-12-29
; PRIOR APPLICATION NUMBER: US 09/382,688
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00590
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; PRIOR FILING DATE: 1997-02-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 18
; TYPE: PRT
; ORGANISM: ORGANISM: Ananas comosus
US-09-750-210A-2

Query Match      81.5%; Score 88; DB 10; Length 18;
Best Local Similarity 88.9%; Pred. No. 1.2e-08;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKN 18
DB 1 VPOSIDWRDYGAVTSVKN 18

RESULT 5
US-09-750-210A-1
; Sequence 1, Application US/09750210A
; Patent No. US20020102253A1
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracy Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0740001
; CURRENT APPLICATION NUMBER: US/09/750,210A
; CURRENT FILING DATE: 2000-12-29
; PRIOR APPLICATION NUMBER: US 09/382,688
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00590
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; PRIOR FILING DATE: 1997-02-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 18
; TYPE: PRT
; ORGANISM: ORGANISM: Ananas comosus
US-09-750-210A-1

Query Match      81.5%; Score 88; DB 10; Length 18;
Best Local Similarity 88.9%; Pred. No. 1.2e-08;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKN 18
DB 1 VPOSIDWRDYGAVTSVKN 18
```

; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 21
; TYPE: PRT
; ORGANISM: ORGANISM: Ananas comosus
US-09-750-210A-1

Query Match 65.7%; Score 71; DB 10; Length 21;
Best Local Similarity 68.4%; Pred. No. 9.3e-06;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VPOSIDMRDYGAVTSVKQ 19
Db 2 LPDSIDMRQKGAVTPEKRN 20

RESULT 6
US-09-462-846-3
; Sequence 3, Application US/09462846
; Patent No. US20020031807A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC381-US
; CURRENT APPLICATION NUMBER: US/09/462,846
; PRIOR FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: PCT/US98/19529
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: EP 97305227.7
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-462-846-3

Query Match 65.7%; Score 71; DB 10; Length 220;
Best Local Similarity 63.2%; Pred. No. 0.00012;
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VPOSIDMRDYGAVTSVKQ 19
Db 9 IPEYVDMRQKGAVTPEKRN 27

RESULT 7
US-09-953-956-9
; Sequence 9, Application US/09953956
; Patent No. US20020072107A1
; GENERAL INFORMATION:
; APPLICANT: Haefliger, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D2D1
; CURRENT APPLICATION NUMBER: US/09/953,956
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: 09/219,441
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 9
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-953-956-9

Query Match 63.9%; Score 69; DB 10; Length 333;
Best Local Similarity 66.7%; Pred. No. 0.00041;
Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 POSIDMRDYGAVTSVKQ 19

Db 115 PRSVDMRKGYVTPVKQ 132

RESULT 8
US-10-114-464-9
; Sequence 9, Application US/10114464
; Patent No. US20020142448A1
; GENERAL INFORMATION:
; APPLICANT: Haefliger, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D5
; CURRENT APPLICATION NUMBER: US/10/114,464
; PRIOR FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 08/553,125
; PRIOR FILING DATE: 1995-11-07
; PRIOR APPLICATION NUMBER: 08/208,007
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 9
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-114-464-9

Query Match 63.9%; Score 69; DB 12; Length 333;
Best Local Similarity 66.7%; Pred. No. 0.00041;
Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 POSIDMRDYGAVTSVKQ 19
Db 115 PRSVDMRKGYVTPVKQ 132

RESULT 9
US-09-529-063-1
; Sequence 1, Application US/09529063
; Patent No. US20020102542A1
; GENERAL INFORMATION:
; APPLICANT: FUKUSHIMA, DAICHI
; APPLICANT: SHIBAYAMA, SHIRO
; APPLICANT: TADA, HIDEAKI
; TITLE OF INVENTION: POLYPEPTIDE, CDNA ENCODING THE POLYPEPTIDE, AND USE OF
; FILE REFERENCE: Q58769
; CURRENT APPLICATION NUMBER: US/09/529,063
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: PCT/JP98/04514
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: JP 9-274674
; PRIOR FILING DATE: 1997-10-07
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-529-063-1

Query Match 62.0%; Score 67; DB 10; Length 334;
Best Local Similarity 63.2%; Pred. No. 0.00088;
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VPOSIDMRDYGAVTSVKQ 19
Db 114 LPRSVDMRKKGYVTPVKQ 132

RESULT 10
US-09-990-064-2
; Sequence 2, Application US/09990064
; Patent No. US20020164765A1

```
; GENERAL INFORMATION:
; APPLICANT: Parvinder K. Punia
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30242
; CURRENT APPLICATION NUMBER: US/09/990,064
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: GB0028462.0
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: fastseq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Mammalia
US-09-990-064-2

Query Match      59.3%; Score 64; DB 9; Length 331;
Best Local Similarity 57.9%; Pred. No. 0.0027;
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKNQ 19
Db 115 LPDSVDWREKGCVTVEKYQ 133

RESULT 11
US-09-953-956-8
; Sequence 8, Application US/09953956
; Patent No. US20020072107A1
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D21
; CURRENT APPLICATION NUMBER: US/09/953,956
; CURRENT FILING DATE: 2001-09-18
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 09/219,441
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-953-956-8

Query Match      59.3%; Score 64; DB 10; Length 331;
Best Local Similarity 57.9%; Pred. No. 0.0027;
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPOSIDWRDYGAVTSVKNQ 19
Db 115 LPDSVDWREKGCVTVEKYQ 133

RESULT 12
US-10-114-464-8
; Sequence 8, Application US/10114464
; Patent No. US20020142448A1
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D5
; CURRENT APPLICATION NUMBER: US/10/114,464
; CURRENT FILING DATE: 2002-04-03
; PRIOR FILING DATE: 1995-11-07
; PRIOR APPLICATION NUMBER: 08/553,125
; PRIOR FILING DATE: 1994-03-08
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-114-464-8

Query Match      52.8%; Score 57; DB 12; Length 329;
Best Local Similarity 66.7%; Pred. No. 0.039;
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 POSIDWRDYGAVTSVKNQ 19
Db 116 PDSIDYRKKGVTVPVKQ 133

RESULT 13
US-09-953-956-7
; Sequence 7, Application US/09953956
; Patent No. US20020072107A1
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D21
; CURRENT APPLICATION NUMBER: US/09/953,956
; CURRENT FILING DATE: 2001-09-18
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 09/219,441
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-953-956-7

Query Match      52.8%; Score 57; DB 10; Length 329;
Best Local Similarity 66.7%; Pred. No. 0.039;
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 POSIDWRDYGAVTSVKNQ 19
Db 116 PDSIDYRKKGVTVPVKQ 133

RESULT 14
US-10-114-464-7
; Sequence 7, Application US/10114464
; Patent No. US20020142448A1
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D5
; CURRENT APPLICATION NUMBER: US/10/114,464
; CURRENT FILING DATE: 2002-04-03
; PRIOR FILING DATE: 08/553,125
; PRIOR FILING DATE: 1995-11-07
; PRIOR APPLICATION NUMBER: 08/208,007
; PRIOR FILING DATE: 1994-03-08
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-114-464-7

Query Match      52.8%; Score 57; DB 12; Length 329;
Best Local Similarity 66.7%; Pred. No. 0.039;
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 POSIDWRDYGAVTSVKNQ 19
Db 116 PDSIDYRKKGVTVPVKQ 133

RESULT 15
```

```

US-09-953-956-2
Sequence 2, Application US/09953956
Patent No. US20020072107a1
GENERAL INFORMATION:
APPLICANT: Hastings, et al.
TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
FILE REFERENCE: P1107D2D1
CURRENT APPLICATION NUMBER: US/09/953,956
CURRENT FILING DATE: 2001-09-18
PRIOR APPLICATION NUMBER: 09/219,441
PRIOR FILING DATE: 1998-12-23
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 329
TYPE: PRT
ORGANISM: Homo sapiens
US-09-953-956-2

```

Query Match	51.9%	Score 56	DB 10	Length 329
Best Local Similarity	61.1%	Pred. No. 0.057		
Matches 1	Conservative	2	Mismatches 5	Indels 0
			Gaps	0

QY	2	PQSIDWRDYGAVTSVKNO	19
		:	
Db	116	PDSVDYRKKGYTTPVKNO	133

Search completed: November 21, 2002, 16:16:25
Job time : 7 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2002, 16:10:04 ; Search time 7 Seconds

(without alignments)
44.747 Million cell updates/sec

Title: US-09-674-738-2

Sequence: 1 VPOSIDWRDSCGATSVKNOG 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100480 seqs, 15661496 residues

Total number of hits satisfying chosen parameters: 100480

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB_PEP:*
- 2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB_PEP:*
- 3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB_PEP:*
- 4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB_PEP:*
- 5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB_PEP:*
- 6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB_PEP:*
- 7: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB_PEP:*
- 8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB_PEP:*
- 9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB_PEP:*
- 10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB_PEP:*
- 11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB_PEP:*
- 12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB_PEP:*
- 13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB_PEP:*
- 14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB_PEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	100.0	20	US-09-750-210A-4	Sequence 4, Appli
2	94	89.5	21	US-09-750-210A-5	Sequence 5, Appli
3	93	88.6	21	US-09-750-210A-3	Sequence 3, Appli
4	79	75.2	18	US-09-750-210A-2	Sequence 2, Appli
5	79	75.2	21	US-09-750-210A-1	Sequence 1, Appli
6	79	75.2	220	US-09-462-846-3	Sequence 3, Appli
7	77	73.3	333	US-09-953-956-9	Sequence 9, Appli
8	77	73.3	333	US-10-114-464-9	Sequence 9, Appli
9	72	68.6	331	US-09-953-956-8	Sequence 8, Appli
10	72	68.6	331	US-10-114-464-8	Sequence 8, Appli
11	71	67.6	331	US-09-990-064-2	Sequence 2, Appli
12	69	63.7	334	US-09-529-063-1	Sequence 1, Appli
13	65	61.9	329	US-09-953-956-7	Sequence 7, Appli
14	65	61.9	329	US-10-114-464-7	Sequence 7, Appli
15	64	61.0	329	US-09-953-956-2	Sequence 2, Appli
16	64	61.0	329	US-10-114-464-2	Sequence 2, Appli
17	64	61.0	361	US-09-925-300-1693	Sequence 1693, Ap
18	57.5	54.8	335	US-09-953-956-10	Sequence 10, Appli
19	57.5	54.8	335	US-10-114-464-10	Sequence 10, Appli

20	57.5	54.8	364	US-09-925-302-478	Sequence 478, App
21	57	54.3	337	US-09-925-301-1021	Sequence 1021, Ap
22	45.5	43.3	165	US-09-864-761-42873	Sequence 42873, A
23	44	41.9	565	US-09-925-301-1302	Sequence 1302, Ap
24	43	41.0	119	US-09-820-893-50	Sequence 50, Appli
25	43	41.0	1247	US-09-815-242-10145	Sequence 10145, A
26	43	41.0	1247	US-09-815-242-13841	Sequence 13841, A
27	42	40.0	185	US-09-893-737-78	Sequence 78, Appli
28	42	40.0	343	US-09-741-669-412	Sequence 412, App
29	42	40.0	574	US-09-782-980-15	Sequence 15, Appli
30	42	40.0	574	US-09-909-743-6	Sequence 6, Appli
31	42	40.0	1165	US-10-016-768-10	Sequence 10, Appli
32	42	40.0	1221	US-10-016-768-11	Sequence 11, Appli
33	41	39.0	1221	US-09-738-973-100	Sequence 100, App
34	41	39.0	127	US-09-738-973-101	Sequence 101, App
35	41	39.0	190	US-09-811-284-227	Sequence 227, App
36	41	39.0	457	US-09-815-242-10370	Sequence 10370, A
37	41	39.0	516	US-09-986-676A-2	Sequence 2, Appli
38	41	39.0	516	US-09-971-611-2	Sequence 110, App
39	40.5	38.6	472	US-09-800-729-130	Sequence 95, Appli
40	40.5	38.6	626	US-09-822-217-1117	Sequence 1117, Ap
41	39.5	37.6	806	US-09-922-217-1118	Sequence 1118, Ap
42	39.5	37.6	806	US-09-729-674-160	Sequence 160, App
43	39	37.1	423	US-09-729-674-160	Sequence 13686, A
44	39	37.1	637	US-09-815-242-13686	Sequence 2, Appli
45	39	37.1	637	US-09-862-007-2	

ALIGNMENTS

RESULT 1
US-09-750-210A-4
Sequence 4, Application US/09750210A
Patent No. US20020102253A1
GENERAL INFORMATION:
APPLICANT: Wyott, Tracy Lehane
APPLICANT: Engwerda, Christian
APPLICANT: Peek, Keith
TITLE OF INVENTION: Component of Bromelain
FILE REFERENCE: 0623.0740001
CURRENT APPLICATION NUMBER: US/09/750,210A
CURRENT FILING DATE: 2000-12-29
PRIOR APPLICATION NUMBER: US-09/382,688
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: PCT/GB98/00590
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: GB 9706119.6
PRIOR FILING DATE: 1997-03-25
PRIOR APPLICATION NUMBER: GB 9704252.7
PRIOR FILING DATE: 1997-02-28
PRIOR APPLICATION NUMBER: GB 9703850.9
PRIOR FILING DATE: 1997-02-25
PRIOR APPLICATION NUMBER: GB 9703827.7
PRIOR FILING DATE: 1997-02-25
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 20
TYPE: PRT
ORGANISM: ORGANISM: Ananas comosus
US-09-750-210A-4

Query Match 100.0%; Score 105; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.8e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSCGATSVKNOG 20
Db 1 VPOSIDWRDSCGATSVKNOG 20

RESULT 2

US-09-750-210A-5
; Sequence 5, Application US/09750210A
; Patent No. US20020102253A1
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracy Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0740001
; CURRENT APPLICATION NUMBER: US/09/750,210A
; CURRENT FILING DATE: 2000-12-29
; PRIOR APPLICATION NUMBER: US 09/382,688
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00590
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; PRIOR FILING DATE: 1997-02-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 20
; TYPE: PRT
; ORGANISM: ORGANISM: Ananas comosus
US-09-750-210A-5

Query Match 89.5%; Score 94; DB 10; Length 20;
Best Local Similarity 90.0%; Pred. No. 1.2e-09;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPQSIDWRDSDGAVTSVKNOG 20
Db 1 VPQSIDWRNYGAVTSVKNOG 20

RESULT 3
US-09-750-210A-3
; Sequence 3, Application US/09750210A
; Patent No. US20020102253A1
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracy Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0740001
; CURRENT APPLICATION NUMBER: US/09/750,210A
; CURRENT FILING DATE: 2000-12-29
; PRIOR APPLICATION NUMBER: US 09/382,688
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00590
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; PRIOR FILING DATE: 1997-02-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 21
; TYPE: PRT
; ORGANISM: ORGANISM: Ananas comosus
US-09-750-210A-3

Query Match 88.6%; Score 93; DB 10; Length 21;

Best Local Similarity 94.7%; Pred. No. 1.9e-09;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 VPQSIDWRDSDGAVTSVKNO 19
Db 2 VPQSIDWRDYGAVTSVKNO 20
RESULT 4
US-09-750-210A-2
; Sequence 2, Application US/09750210A
; Patent No. US20020102253A1
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracy Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0740001
; CURRENT APPLICATION NUMBER: US/09/750,210A
; CURRENT FILING DATE: 2000-12-29
; PRIOR APPLICATION NUMBER: US 09/382,688
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00590
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; PRIOR FILING DATE: 1997-02-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 18
; TYPE: PRT
; ORGANISM: ORGANISM: Ananas comosus
US-09-750-210A-2

Query Match 75.2%; Score 79; DB 10; Length 18;
Best Local Similarity 83.3%; Pred. No. 3.3e-07;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPQSIDWRDSDGAVTSVKN 18
Db 1 VPQSIDWRDYGAVNEVKN 18

RESULT 5
US-09-750-210A-1
; Sequence 1, Application US/09750210A
; Patent No. US20020102253A1
; GENERAL INFORMATION:
; APPLICANT: Mynott, Tracy Lehanne
; APPLICANT: Engwerda, Christian
; APPLICANT: Peek, Keith
; TITLE OF INVENTION: Component of Bromelain
; FILE REFERENCE: 0623.0740001
; CURRENT APPLICATION NUMBER: US/09/750,210A
; CURRENT FILING DATE: 2000-12-29
; PRIOR APPLICATION NUMBER: US 09/382,688
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/GB98/00590
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: GB 9706119.6
; PRIOR FILING DATE: 1997-03-25
; PRIOR APPLICATION NUMBER: GB 9704252.7
; PRIOR FILING DATE: 1997-02-28
; PRIOR APPLICATION NUMBER: GB 9703850.9
; PRIOR FILING DATE: 1997-02-25
; PRIOR APPLICATION NUMBER: GB 9703827.7
; PRIOR FILING DATE: 1997-02-25

```

; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 21
; TYPE: PRT
; ORGANISM: ORGANISM: Ananas comosus
US-09-750-210A-1

```

```

Query Match 75.2%; Score 79; DB 10; Length 21;
Best Local Similarity 70.0%; Pred. No. 3.9e-07;
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 1 VPOSIDWRDSCGAVTSYKNOG 20
:|||||:|||||:|||||:
Db 2 LPDSIDWRQKGAVTGVKNOG 21

```

```

RESULT 6
US-09-462-846-3
; Sequence 3, Application US/09462846
; Patent No. US20020031807A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC381-US
; CURRENT APPLICATION NUMBER: US/09/462,846
; PRIOR FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: PCT/US98/19529
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: EP 97305227.7
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-462-846-3

```

```

Query Match 75.2%; Score 79; DB 10; Length 220;
Best Local Similarity 65.0%; Pred. No. 5.5e-06;
Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 1 VPOSIDWRDSCGAVTSYKNOG 20
:|||||:|||||:|||||:
Db 9 IPEYDWRQKGAVTGVKNOG 28

```

```

RESULT 7
US-09-953-956-9
; Sequence 9, Application US/09953956
; Patent No. US20020072107A1
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D2D1
; CURRENT APPLICATION NUMBER: US/09/953,956
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: 09/219,441
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 9
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-953-956-9

```

```

Query Match 73.3%; Score 77; DB 10; Length 333;
Best Local Similarity 68.4%; Pred. No. 1.9e-05;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 2 POSIDWRDSCGAVTSYKNOG 20

```

```

Db 115 PRSDVWRKRGVTVPVKNOG 133
:|||||:|||||:|||||:

```

```

RESULT 8
US-10-114-464-9
; Sequence 9, Application US/10114464
; Patent No. US20020142448A1
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D5
; CURRENT APPLICATION NUMBER: US/10/114,464
; PRIOR FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 08/553,125
; PRIOR FILING DATE: 1995-11-07
; PRIOR APPLICATION NUMBER: 08/208,007
; PRIOR FILING DATE: 1994-03-08
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 9
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-114-464-9

```

```

Query Match 73.3%; Score 77; DB 12; Length 333;
Best Local Similarity 68.4%; Pred. No. 1.9e-05;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 2 POSIDWRDSCGAVTSYKNOG 20
:|||||:|||||:|||||:
Db 115 PRSDVWRKRGVTVPVKNOG 133

```

```

RESULT 9
US-09-953-956-8
; Sequence 8, Application US/09953956
; Patent No. US20020072107A1
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D2D1
; CURRENT APPLICATION NUMBER: US/09/953,956
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: 09/219,441
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 8
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-953-956-8

```

```

Query Match 68.6%; Score 72; DB 10; Length 331;
Best Local Similarity 60.0%; Pred. No. 0.00013;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 VPOSIDWRDSCGAVTSYKNOG 20
:|||||:|||||:|||||:
Db 115 LPDSVWRKRGVTVPVKNOG 134

```

```

RESULT 10
US-10-114-464-8
; Sequence 8, Application US/10114464
; Patent No. US20020142448A1
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D5
; CURRENT APPLICATION NUMBER: US/10/114,464

```

```
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 08/553,125
; PRIOR FILING DATE: 1995-11-07
; PRIOR APPLICATION NUMBER: 08/208,007
; PRIOR FILING DATE: 1994-03-08
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-114-464-8

Query Match      68.6%; Score 72; DB 12; Length 331;
Best Local Similarity 60.0%; Pred. No. 0.00013;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSGAVTSVKNOG 20
Db 115 LPDSVDWRKGCVTVEKYQG 134

RESULT 11
US-09-990-064-2
; Sequence 2, Application US/09990064
; Patent No. US20020164765A1
; GENERAL INFORMATION:
; APPLICANT: Parvinder K. Punia
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30242
; CURRENT APPLICATION NUMBER: US/09/990,064
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: GB0028462.0
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Mammalia
US-09-990-064-2

Query Match      67.6%; Score 71; DB 9; Length 331;
Best Local Similarity 60.0%; Pred. No. 0.00019;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSGAVTSVKNOG 20
Db 115 LPDSVDWRKGCVTVEKYQG 134

RESULT 12
US-09-529-063-1
; Sequence 1, Application US/09529063
; Patent No. US20020102542A1
; GENERAL INFORMATION:
; APPLICANT: FUKUSHIMA, DAIKICHI
; APPLICANT: SHIBAYAMA, SHIRO
; APPLICANT: TADA, HIDEAKI
; TITLE OF INVENTION: POLYPEPTIDE, cDNA ENCODING THE POLYPEPTIDE, AND USE OF
; FILE REFERENCE: Q58769
; CURRENT APPLICATION NUMBER: US/09/529,063
; CURRENT FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: PCT/JP98/04514
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: JP 9-274674
; PRIOR FILING DATE: 1997-10-07
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 334
; TYPE: PRT

; ORGANISM: Homo sapiens
US-09-529-063-1

Query Match      65.7%; Score 69; DB 10; Length 334;
Best Local Similarity 63.2%; Pred. No. 0.0004;
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 VPOSIDWRDSGAVTSVKNOG 19
Db 114 LPKSVDWRKKGVTVPVKNOG 132

RESULT 13
US-09-953-956-7
; Sequence 7, Application US/09953956
; Patent No. US20020072107A1
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D2D1
; CURRENT APPLICATION NUMBER: US/09/953,956
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: 09/219,441
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-953-956-7

Query Match      61.9%; Score 65; DB 10; Length 329;
Best Local Similarity 68.4%; Pred. No. 0.0018;
Matches 13; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 POSIDWRDSGAVTSVKNOG 20
Db 116 PDSIDYRKKGVTVPVKNOG 134

RESULT 14
US-10-114-464-7
; Sequence 7, Application US/10114464
; Patent No. US20020142448A1
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D5
; CURRENT APPLICATION NUMBER: US/10/114,464
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 08/553,125
; PRIOR FILING DATE: 1995-11-07
; PRIOR APPLICATION NUMBER: 08/208,007
; PRIOR FILING DATE: 1994-03-08
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-114-464-7

Query Match      61.9%; Score 65; DB 12; Length 329;
Best Local Similarity 68.4%; Pred. No. 0.0018;
Matches 13; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 POSIDWRDSGAVTSVKNOG 20
Db 116 PDSIDYRKKGVTVPVKNOG 134

RESULT 15
```

```

US-09-953-956-2
Sequence 2, Application US/09953956
Patent No. US20020072107A1
GENERAL INFORMATION:
APPLICANT: Haeftings, et al.
FILE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
FILE REFERENCE: P1107D201
CURRENT APPLICATION NUMBER: US/09/953,956
PRIORITY FILING DATE: 2001-09-18
PRIOR APPLICATION NUMBER: 09/219,441
PRIOR FILING DATE: 1998-12-23
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 329
TYPE: PR1
ORGANISM: Homo sapiens
US-09-953-956-2

```

Query Match	61.0%;	Score 64;	DB 10;	Length 329;
Best Local Similarity	63.2%;	Pred. No. 0.0027;		
Matches 12;	Conservative 2;	Mismatches 5;	Indels 0;	Gaps 0;

```
Qy      2  P Q S I D W R D S G A V T S V K N Q G 20
          | | : | : | | | | | | |
Db      116 P D S V D Y R K K G Y V T P V K N Q G 134
```

Search completed: November 21, 2002, 16:16:26
Job time : 8 secs

47x2